UNIVERSIDAD DEL PACÍFICO | Derecho
Facultad de Derecho

EL IMPACTO DEL WEB SCRAPING EN LA PRIVACIDAD: ¿HACIA UNA NUEVA REGULACIÓN? UN ANÁLISIS DESDE LA REGULACIÓN DE PROTECCIÓN DE DATOS PERSONALES

Tesis presentada para optar al Título Profesional de Abogada

Presentado por:
Ana Lucia Taboada Cardenas
Adara Lucia Palomino Gonzáles

Asesor: Julio Álvaro Castro Lora
0009-0005-8800-0033

Lima, agosto 2024

REPORTE DE EVALUACIÓN DEL SISTEMA ANTIPLAGIO

FACULTAD DE DERECHO

A través del presente documento la Facultad de Derecho deja constancia de que la Tesis titulada ***"El impacto del Web Scraping en la privacidad: ¿Hacia una nueva regulación? Un análisis desde la regulación de protección de datos personales"*** presentado por la Srta. ANA LUCÍA TABOADA CÁRDENAS, con DNI No. 72808902, y la Srta. ADARA LUCÍA PALOMINO GONZÁLES, con DNI No. 71404363, para optar el Título Profesional de Abogada, fue sometido al análisis del sistema antiplagio Turnitin el 8 de agosto del año 2024; obteniendo el siguiente resultado:

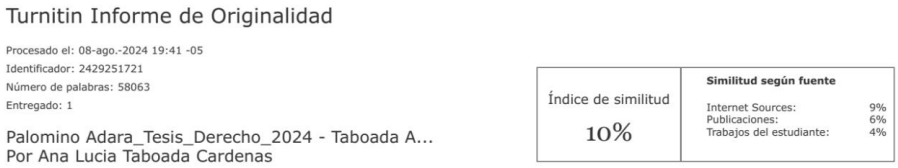
Palomino Adara_Tesis_Derecho_2024 - Taboada Ana Lucia_Tesis_Derecho_2024.docx

INFORME DE ORIGINALIDAD

10%	9%	6%	4%
INDICE DE SIMILITUD	FUENTES DE INTERNET	PUBLICACIONES	TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	hdl.handle.net Fuente de Internet	2%
2	www.informatica-juridica.com Fuente de Internet	1%
3	repositorio.ulima.edu.pe Fuente de Internet	1%
4	Submitted to Universidad de Las Palmas de Gran Canaria Trabajo del estudiante	1%
5	www.cidesi.com Fuente de Internet	1%



Turnitin Informe de Originalidad

Procesado el: 08-ago.-2024 19:41 -05
Identificador: 2429251721
Número de palabras: 58063
Entregado: 1

Palomino Adara_Tesis_Derecho_2024 - Taboada A...
Por Ana Lucia Taboada Cardenas

Índice de similitud	Similitud según fuente	
10%	Internet Sources:	9%
	Publicaciones:	6%
	Trabajos del estudiante:	4%

De acuerdo con la política vigente, el porcentaje obtenido de similitud con otras fuentes está dentro de los márgenes permitidos.

Se emite el presente documento para los fines estipulados en el Reglamento de Grados y Títulos de la Facultad al que pertenece el interesado.

Lima, 5 de septiembre de 2024

Julio Álvaro Castro Lora
Asesor
Código Orcid: 0009-0005-8800-0033

A mis padres y a mis abuelos, que me supieron guiar y aconsejar en los momentos adversos. A mi hermano, por acompañarme desde el inicio de mi etapa universitaria y a mi ángel quien desde el cielo me ha motivado a crecer de manera personal y profesional.

Adara Palomino

A mis padres, Esther y Luis, y a mis abuelitos, sin ustedes no sería lo que soy. A Candy, quien ahora me acompaña desde el cielo, y a Micky, que llegaste cuando inició mi viaje universitario. A mis jefes, compañeros y amigos, quienes me han motivado a cuestionar y buscar nuevos caminos para lograr objetivos innovadores y disruptivos. A Álvaro, nuestro asesor, por incentivarnos a crear nuevas realidades en el Derecho.

Ana Lucia Taboada

Privacy is not about something to hide. Privacy is about something to protect. And that's who you are. That's what you believe in. That's who you want to become.

Edward Snowden

"*Scraping" does not have to be a dirty word.*

Ethical Web Data Collection Initiative

RESUMEN

El *Web Scraping* es una técnica usualmente utilizada para extraer información de personas naturales, incluyendo sus datos personales. *Prima facie,* el *Web Scraping* puede resultar inofensivo al permitir crear estadísticas y/o cálculos que permitan la toma de decisiones; sin embargo, habitualmente se extraen datos personales incumpliendo con los principios fundamentales de la protección de datos personales, resaltando los principios de consentimiento e información.

En ese sentido, la presente tesis busca determinar si es necesario regular específicamente el fenómeno del *Web Scraping* dado su impacto en la privacidad de los usuarios y sus implicancias dentro de la regulación peruana de protección de datos personales. Por ello, realizaremos un análisis metodológico y técnico sobre el *Web Scraping* a fin de entender su naturaleza. En esa línea, se estudiará la naturaleza de un dato personal a nivel constitucional y regulatorio, explicando los tipos y/o clasificaciones existentes. Seguidamente, se revisará las medidas optadas por regulación extranjera, especialmente la europea. Finalmente, de lo analizado, se expondrá la respuesta a nuestra interrogante que nos llevó a redactar la presente tesis.

El presente trabajo desarrollará un método descriptivo en tanto explica las definiciones pertinentes, así como la mención de la normativa y jurisprudencia aplicables; técnico al profundizar la naturaleza del *Web Scraping* presentando sus alcances y ejecución mediante una simulación práctica; y analítico al realizar un estudio profundo de la teoría y práctica que nos permite responder nuestra hipótesis.

Palabras clave: Web scraping, extracción, datos personales, consentimiento, tratamiento.

ABSTRACT

Web Scraping is a technique usually used to extract information from natural persons, including their personal data. *Prima facie*, *Web Scraping* may be harmless as it allows the creation of statistics and/or estimates that allow decision making; however, personal data is usually extracted violating the principles of personal data protection, emphasizing the principles of consent and information.

This thesis seeks to identify whether it is necessary to specifically regulate the phenomenon of *Web Scraping* given its impact on users' privacy and its implications with personal data protection regulations. For this purpose, we will conduct a methodological and technical analysis of *Web Scraping* in order to understand its nature. We will study the nature of personal data at constitutional and regulatory level, explaining the existing types and/or classifications. Then, we will review the measures chosen by foreign regulations, especially the EU one. Finally, from what has been analyzed, the answer to our question that led us to write this thesis will be presented.

This study will develop a descriptive method in explaining the pertinent definitions, as well as the mention of the applicable regulations and jurisprudence; technical by deepening the nature of *Web Scraping* presenting its scope and execution through a practical simulation; and an analytical method in carrying out an in-depth study of the theory and practice that will allow us to response our hypothesis.

Key words: Web scraping, extraction, personal data, consent, personal data processing.

TABLA DE CONTENIDO

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

INTRODUCCIÓN

El Internet se ha constituido como uno de los espacios más grandes de compartimiento de información y, con ello, el intercambio inevitable, incesante y continuo de datos personales. Tan es así que los datos personales representan un activo importante que se compra y vende en un mercado activo multimillonario. Bajo esa línea, la forma de recolección de los datos personales es un elemento clave para el mundo del intercambio de información. Es evidente que este ha ido cambiando a través del tiempo, en tanto se han originado innovaciones disruptivas en la forma que se producen, gestionan, analizan, almacenan y utilizan los datos (Christensen, 1997).

Como parte de este fenómeno de extracción de datos en la era digital, es que aparece la figura del *Web Scraping*, el cual, a grandes rasgos, constituye una técnica de extracción automatizada de datos de páginas web y/o plataformas digitales. Debido a que se extrae información de manera masiva de estas, entonces queda al descubierto el riesgo potencial de que se puedan extraer datos personales ilegalmente por incumplir con las disposiciones en materia de protección de datos personales vigente.

El pasado junio de 2023, la Corte Federal de San Francisco (PM v. OpenAI LP, 2023) recibió una demanda en contra de OpenAI y Microsoft Corporation por extraer datos personales, mediante el *Web Scraping*, a cientos de millones de usuarios de Internet, incluidos niños, sin el debido consentimiento. Estos percances no son recientes, pues también en el pasado mayo de 2017, LinkedIn demandó a hiQ por extraer las imágenes, a través del *Web Scraping*, de los perfiles públicos de la plataforma (hiQ Labs, Inc. v. LinkedIn Corp, 2019). Es evidente que conforme las nuevas tecnologías sigan sumando nuevos elementos y alcances para sus productos finales es que el *Web Scraping* toma mayor trascendencia y, con ello, el riesgo latente de un ilegal y fraudulento tratamiento de datos personales.

En ese sentido, el objeto del presente trabajo es determinar si es que se debe regular el uso de la técnica del *Web Scraping* dado sus implicancias jurídicas que impactan, principalmente, en la legislación de protección de datos personales peruana.

Para ello, en el Capítulo I se explica la naturaleza del *Web Scraping* analizando el entorno que lo rodea, los actores participantes y los procesos implicados para que se pueda generar, como tal, la extracción de datos. Para mayor entendimiento, se realiza una simulación de *Web*

Scraping explicado paso a paso y de manera gráfica cómo un usuario de internet puede ejecutar dicha técnica. Así también, se responde una pregunta clave dado los grandes avances de la inteligencia artificial y que, por tanto, permiten la simplificación del *Web Scraping* en el entorno digital.

Seguidamente, el Capítulo II realiza una revisión de lo que se entiende por dato como raíz elemental para la formación del término dato personal. Así, se analiza el concepto de datos personales y su clasificación, incluyendo la legislación local e internacional, aunado a los documentos internacionales que han tratado de definirlo junto con sus alcances. Se incluye a la doctrina autorizada que permite dilucidar con mayor precisión la clasificación de los datos personales. Sumado a ello, resulta inconcebible no mencionar las implicancias constitucionales protegidas por el derecho a la protección de datos personales, siendo que se hace mención al derecho a la intimidad personal y familiar, el derecho a la autodeterminación informativa y el proceso de *Habeas Data*. Finalizando dicho capítulo, se realiza un análisis general de la regulación peruana de protección de datos personales que está conformada por la Ley No. 29733, Ley de Protección de Datos Personales (2011), su Reglamento (2013) y directivas aplicables.

En esa línea, el Capítulo III ofrece una amplia variedad de medidas regulatorias extranjeras, tales como países de la Unión Europea y América, los cuales nos permitirán conocer, en base a su regulación, cuáles son las medidas optadas en contra de la extracción de datos personales, incluyendo el *Web Scraping*. También se suma la jurisprudencia que nos permitirá analizar a detalle el razonamiento del legislador extranjero.

El Capítulo IV constituye el análisis de los matices y principios de la regulación peruana de protección de datos personales aplicada a la técnica del *Web Scraping*. Es aquí donde, al amparo de lo mencionado y revisado en los capítulos anteriores, se da respuesta al objetivo de la presente tesis. Sin ánimos de adelantar dicho análisis, consideramos que no es necesario regular *per se* el *Web Scraping*, en tanto la normativa peruana es aplicable a dicha técnica de extracción de datos. En base a ello, se detallan los aspectos a considerar para su aplicación, así como una propuesta práctica y funcional de acuerdo a su naturaleza.

CAPÍTULO I. NATURALEZA Y FUNCIONAMIENTO DEL WEB SCRAPING

1.1 El entorno del *Web Scraping*

En los últimos tiempos, el Internet se ha considerado una red global de información interconectada, escalable y expansiva que permite acceder a gran cantidad de información y recursos al alcance de un solo *click*. El Internet es una infraestructura de comunicaciones que permite la interconexión de diferentes redes, por lo que se le considera como una red de redes (Internet Society, 2020). Es una a red informática mundial de uso público, que proporciona acceso a una serie de servicios de comunicación, incluyendo la web, y que transporta correo electrónico, noticias, entretención y archivos de datos (Lineamientos para la Formulación del Plan de Gobierno Digital, 2022). Esto es gracias a los distintos servidores y redes interconectadas en todo el mundo que almacenan y procesan una amplia gama de datos, siendo además que el éxito del Internet se debe, en gran parte, a su modelo único: propiedad global compartida, desarrollo de estándares abiertos y transparentes, además de procesos de libre acceso para el para el desarrollo tecnológico y político (Internet Society, s.f.).

Desde sus inicios, el Internet nació con el propósito de transportar información protegiendo los datos que se podían emitir, a pesar de que los centros físicos sean destruidos (Anzures Gurría, 2020). Ello ocurrió a través de la creación del proyecto militar denominado ARPAnet (*Advanced Research Projects Agency Network*), una red informática ejecutada durante la Guerra Fría como un medio para enviar datos militares de Estados Unidos y comunicar a sus principales grupos de investigación. Gracias a dicho proyecto, el Internet comienza con un despliegue abismal, siendo que, en 1973, se crean el Protocolo de Control de Transmisión (TCP[1]) y el Protocolo de Internet (IP[2]), elementos claves que permiten el desarrollo del Internet a un nivel más escalable.

Posteriormente, en el siglo XXI, el Internet es testigo de una transformación a la Web 2.0, en donde el usuario genera y transmite información, dejando de lado la única función de recibir

[1] Cuenta con la responsabilidad de garantizar que los datos sean entregados de manera confiable y ordenada entre aplicaciones en diferentes hosts. A su vez, utiliza un mecanismo de control para evitar la sobrecarga de la red y garantizar un flujo de datos adecuado. Recuperado de: https://www.semanticscholar.org/paper/Dom%C3%B3tica-mediante-la-transmisi%C3%B3n-de-datos-digitales-Valer-Rodr%C3%ADguez/83ba6577e54d70b40e33feccd17c33562b4fade1.

[2] El protocolo de internet tiene como función elemental establecer el direccionamiento en una red. Sin ella resulta imposible la comunicación entre los dispositivos, ya que esta tiene la información para navegar por redes. Recuperado de: https://www.semanticscholar.org/paper/Gu%C3%ADa-pr%C3%A1ctica-de-simulaci%C3%B3n-e-implementaci%C3%B3n-del-de-Valderrama-Riveros/9549d8686f34a717c16aa8f38db0edee0c6614ef.

data. Esto ha sido el punto de partida para lo que después se ha denominado la Cuarta Revolución Industrial, un fenómeno referido a la transformación en curso de las industrias y las sociedades mediante la integración de las tecnologías digitales, la automatización y los procesos basados en datos (Lee et al., 2018).

A la fecha, el Perú reconoce la importancia de contar con un internet seguro, libre de peligros y amenazas para todos los ciudadanos digitales, con especial énfasis a las niñas, niños y adolescentes. Es por ello que, recientemente, se ha creado la Alianza Nacional por una Internet Segura ("Alianza") (Secretaría de Gobierno y Transformación Digital, 2024), en el cual participan distintos actores del medio: entidades públicas, privados, comunidad técnica, academia y sociedades civiles con el propósito de crear, diseñar, ejecutar e implementar iniciativas que promuevan el uso seguro del Internet.

Esta Alianza nace en línea con el propósito de la Ley No. 30254, Ley de Promoción para el Uso Seguro y Responsable de las Tecnologías de la Información y Comunicaciones por Niños, Niñas y Adolescentes (2014), por el cual establece la creación de una comisión encargada de velar y proponer lineamientos aplicables a la seguridad de los menores y adolescentes peruanos. De ese modo, los miembros de la Alianza contribuirían de la mano con el comité para la elaboración de políticas e intervenciones a favor de un espacio seguro en Internet, siendo además un aspecto clave para el avance del ecosistema digital del país.

Sin lugar a dudas, la naturaleza del internet ha logrado que sea considerada compleja y, en gran manera, multifacética, en cuyo caso se distingue por contar con ciertas características (Internet Society, s.f.):

- Medio para la colaboración y la interacción: el Internet permite la conexión entre las personas y los ordenadores sin importar la ubicación geográfica permitiendo que se generen, por ejemplo, distintas transacciones (tales como el comercio electrónico) y la comunicación informática.
- Infraestructura de la información: contar con una gran cantidad de datos e información en un solo lugar unificado y estructurado hace que el Internet sea denominado como la Infraestructura de Información Nacional (o Global, o Galáctica).

- Es abierta (*open network*): cada usuario del Internet es libre de crear y compartir su propio contenido en el espacio virtual, generando el intercambio de información y conocimiento. (Istituto di Studi Giuridici Internazionali , 2010).

Es así como el Internet viene experimentando una metamorfosis constante y latente. Sin embargo, cabe preguntarnos, ¿cuáles son los elementos o actores clave que permiten su funcionamiento? Según Walker, el funcionamiento del Internet desde la perspectiva técnica se basa en los siguientes elementos clave (Chapman and Hall/CRC, 2012):

- Direcciones de internet ("*Internet Addresses*"): Cada dispositivo cuenta con una dirección IP (Protocolo de Internet) única, el cual es un identificador numérico que permite que se pueda comunicar con otros dispositivos. Las direcciones IP son designadas por los proveedores de servicios de Internet (ISP). Al respecto, se cuenta con un debate si es que las direcciones constituyen datos personales, pues pueden ser estáticas (permanecen al mismo ordenador) o dinámicas (temporales, en donde para cada conexión a internet cambia la dirección IP). Esta última no basta por sí sola para identificar a un usuario; sin embargo, puede hacerlo si combina la dirección IP dinámica con otros datos adicionales que obren en poder del ISP (*Internet Service Provider*).
- Protocolos ("*Protocol Stacks*"): El Internet se basa en un conjunto de protocolos que definen cómo se transmiten y reciben los datos entre los dispositivos. Esta dinámica incluye la transmisión de información en pequeños paquetes, cuya naturaleza son pequeñas unidades de información que son enviadas y recibidas por los dispositivos. Estos pueden ser TCP/IP, HTTP y SMTP.
- Protocolo de Control de Transmisión - TCP ("*Transmission Control Protocol*"): Permite la transmisión fiable y ordenada entre los dispositivos de Internet, garantizando precisión y eficacia inclusive en congestiones en la red.
- Infraestructura de internet ("*Internet Infraestructure*"): Incluye los componentes físicos y lógicos, tales como los centros de datos, cables de fibra óptica y protocolos de red. El mantenimiento de la infraestructura se encuentra a cargo de los proveedores de servicios de Internet (ISP), operadores de red y organismos gubernamentales.
- Jerarquía de enrutamiento ("*The Internet Routing Hierarchy*"): Se trata de la forma en que los datos se enrutan (transmiten) entre los dispositivos en Internet, los cuales a su vez se basan en la dirección IP.

- Nombres de dominio ("*Domain Names*"): son aquellos nombres legibles que se utilizan para identificar a los sitios web. Los nombres de dominio se traducen en direcciones IP a través de un proceso denominado resolución de direcciones, el cual es ejecutado por el Sistema de Nombres de Dominio (Domain Name System - DNS).
- Proveedores de servicios de internet ("*Internet Service Providers - ISP*"): custodian el acceso a la totalidad de los contenidos en Internet y, como tales, tienen amplio poder de control sobre qué información reciben y comunican sus suscriptores (Gaspar et al., 2012). En sí, es el ente encargado de poner a disposición de sus usuarios un espacio de memoria en su servidor (aloja contenidos) o también de disponer una parte de su espacio para alojar páginas web de terceros (Gómez Apac Jorge Baeza, M. G. R., et al.,2021). De acuerdo a la normativa peruana, también son denominados proveedores de servicios digitales, entendidos como toda aquella entidad pública o privada que sea responsable por el diseño, prestación y/o acceso a servicios digitales en el territorio nacional, sin perjuicio de su localización geográfica (D.U. No. 007-2020-PCM, 2020). En la actualidad, existe una discusión vigente si es que los ISP pueden llegar a ser responsables por el contenido alojado en sus servicios de titularidad de sus usuarios, pues puede infringir la normativa de protección de datos personales.
- Proveedor de alojamiento ("*Host Service Provider*" o "*Hosting*"): además de brindar un servicio de almacenamiento como tal, también brinda un servicio de mantenimiento de contenidos en su propio servidor con el fin de que los usuarios tengan la opción de conectarse a Internet por un ISP (D.U. No. 007-2020-PCM, 2020).

Asimismo, existen otros actores que confluyen para el adecuado funcionamiento del Internet. Los **navegadores** ("*web browser"* o "*browser"*), tales como Chrome, Firefox y Safari, son aplicaciones utilizadas para navegar por múltiples páginas web y dominios (Vassio et al., 2018) y que a su vez permiten la interacción y el consumo de información proporcionada por diversas páginas disponibles en la web (Silva, Feitosa y Garcia, 2017).

Estos propios navegadores son conscientes de la frecuencia del *Web Scraping* por parte de distintos usuarios de Internet, por lo que, a la fecha, estos publican sus propios *drivers*, que son softwares que funcionan como intermediarios entre un programa y un *website*, permitiendo la interacción entre ambos y recreando el navegador desde un ambiente de programación. Para utilizar estos *drivers*, los usuarios deberán aceptar los términos y condiciones que incluyen las

licencias de uso para el cliente que utiliza el software. De ese modo, el *driver* permite acceder y extraer información de la *website* de forma automatizada. Por ejemplo, este es el caso de Gecko, el motor del navegador web de Firefox que actualmente es gestionado por Mozilla.

En esa línea, los navegadores pueden ejecutar o solicitar la ejecución de los siguientes lenguajes:

- Lenguajes de marcado: considerados lenguajes que permiten definir la estructura y el formato de un documento, en cuyo caso utilizan etiquetas o marcas para indicar cómo debe mostrarse el contenido del mismo (Sierra et al., 2005). Algunos ejemplos de lenguajes de marcado son: HTML (*HyperText Markup Language*), XML (*Extensible Markup Language)* y DSMLs (*Domain-specific markup languages*). Un ejemplo de escritura HTML sería el siguiente:

Figura 1.

Explicación lenguajes de marcado

Fuente: Elaboración propia

En dicho ejemplo se está creando un elemento “h1” del tipo título 1 que contendrá el texto “Hola Mundo”. A continuación, proyectamos otro ejemplo de cómo se entrelazan el HTML con el navegador:

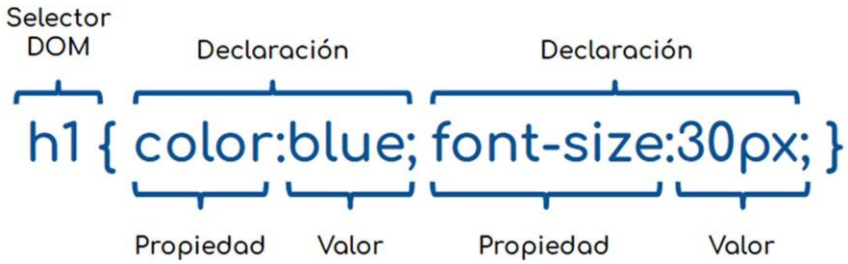
Figura 2.
Explicación del lenguaje HTML y Navegador

Fuente: Elaboración propia

- Lenguajes de estilos: se trata del conjunto de lenguajes informáticos utilizados para definir la presentación y estilo del documento web. Estos lenguajes permiten elegir aspectos como el tamaño del texto, tipografía y las posiciones de los elementos, junto con los colores, de la *website.* Algunos de los más conocidos son CSS (*Cascading Style Sheets*), SASS (*Syntactically Awesome Style Sheets*) y LESS (*Leaner Style Sheets*).

El más conocido es CSS y un ejemplo de su escritura sería el siguiente:

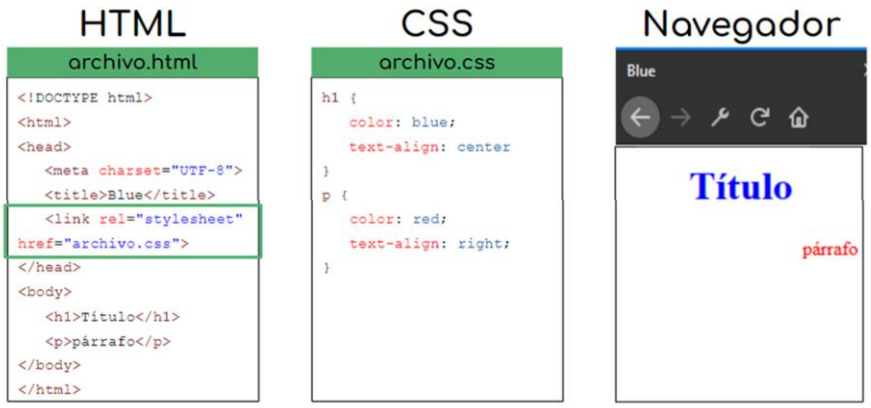
Figura 3.
Explicación de la escritura CSS



Fuente: Elaboración propia

En dicho ejemplo se está ordenando que todos los elementos “h1” del tipo título 1 deben tener un color azul con tamaño de fuente 30 píxeles. A continuación, proyectamos un ejemplo de cómo se entrelazan el CSS con el HTML y navegador:

Figura 4.
Explicación de la sinergia entre el HTML, CSS y Navegador



Fuente: Elaboración propia

- Lenguajes de programación: son aquellos que mediante una secuencia de instrucciones indican a los programaos cómo deben responder ante interacciones en su entorno. Por ello, logran que las acciones que se quieran alcanzar sean posibles (Ramsey, 2022). Estos lenguajes se clasifican según su paradigma de programación; aunque algunos no encajan en un solo paradigma por ser más completos. Sin perjuicio de ello, la clasificación inicial sirve para entender la diversidad y funcionalidad de distintos lenguajes, dividiéndolos en los siguientes:
    I) Procedimentales (*procedural programming languages - "PP"*): sigue un planteamiento lineal para resolver problemas, es decir, un orden específico (Anureev, 2015).
    II) Usualmente son utilizados para procesar información, tales como bases de datos. Suelen ser conocidos por su eficiencia y simplicidad (Ishida, Sasaki, & Fukuhara, 1991). Algunos ejemplos de PP son: C, Pascal y Fortran.
    III) Orientados a objetos (*object-oriented programming languages – "OO"*): es un tipo de lenguaje que utiliza objetos para representar datos y comportamientos (Belyakova,2016). Son conocidos por su flexibilidad y modularidad, lo cual permite a su vez que sean idóneos para el desarrollo de softwares a gran escala. Algunos ejemplos de lenguajes de programación orientados a objetos son Java, Python y C#.
    IV) Funcionales (*functional programming languages*): permiten que los programas se construyan aplicando y componiendo funciones con datos inmutables. Se

centran en la evaluación de expresiones y la transformación de datos, por lo cual ofrecen ventajas interesantes como la mejora de la legibilidad del código, la modularidad y la facilidad de paralelización (Ramsey, 2022). Además, son utilizados en ámbitos como el tratamiento de datos y sistemas distribuidos. Los programas más conocidos son Haskell, Lisp y Erlang.

V) Lógicos*:* En lugar de detallar los pasos o procedimientos, los lenguajes lógicos buscan que el intérprete obtenga soluciones que cumplan con las reglas y condiciones establecidas. Se resalta su deducción lógica, entre los cuales resalta Prolog.

Asimismo, existe otra clasificación para los lenguajes de programación, conforme se detalla a continuación:

- Compilado: son aquellos que requieren de un proceso de compilación para generar un archivo que sea ejecutable para el ordenador. Esto implica que el código fuente bajo el lenguaje de programación se traduzca al código del ordenado[3] (Truong & Hanrahan, 2019) y, con ello, se encuentre en su idioma. Algunos de los ejemplos de programación compilados son C, C++, Fotran y Ada.
- Interpretado: en este caso, los lenguajes de programación cuentan con un intérprete, por lo que no necesitan de un proceso de compilación (Truong & Hanrahan, 2019). Algunos de los ejemplos son Python, Ruby y JavaScript.
- Intermedio: los lenguajes de programación cuentan con un paso intermedio, en tanto luego de escribir el código fuente y compilarlo, obtiene un "Bytecode". Generalmente se utiliza este lenguaje para poder ejecutar el código en cualquier sistema operativo. Ejemplo de este tipo de lenguaje de programación es el Java, Kotlin y Scala.

- Lenguajes de base de datos: o SQL (*Structured Query Language*) es un lenguaje utilizado para definir la estructura de la base de datos, así como su creación y manipulación (Valverde, 2019). Este lenguaje tiene dialectos o variantes, los cuales dependen del motor de base de datos con el que se esté trabajando. Los más conocidos

[3] Lenny Truong and Pat Hanrahan. A Golden Age of Hardware Description Languages: Applying Programming Language Techniques to Improve Design Productivity. In 3rd Summit on Advances in Programming Languages (SNAPL 2019). Leibniz International Proceedings in Informatics (LIPIcs), Volume 136, pp. 7:1-7:21, Schloss Dagstuhl - Leibniz-Zentrum für Informatik (2019). https://doi.org/10.4230/LIPIcs.SNAPL.2019.7

son T- SQL (Microsoft SQL Server), MySQL, PLSQL (Oracle) y PL/pgSOL (PostgreSQL).

En el mundo del desarrollo web, los lenguajes más usados son: JavaScript (lenguaje de programación), HTML (lenguaje de marcado) y CSS (lenguaje de estilo), ya que, combinados, posibilitan la creación y la ejecución de las páginas web que actualmente conocemos en el Internet. No obstante, es común encontrar páginas web con diversos lenguajes de las diferentes categorías mencionadas. Esto se debe a que, en un entorno cada vez más competitivo y exigente, la necesidad de utilizar una conexión más estrecha entre distintos lenguajes se ha vuelto más frecuente.

De la misma forma en la que incrementa la complejidad de los lenguajes, los desarrolladores perfeccionan los entornos de desarrollo en los que operan. Estos últimos se refieren a herramientas de software que proporcionan un entorno integral para la creación de un nuevo software. Incluyen un editor de código, un intérprete y un depurador, siendo que son utilizados para la escritura de pequeños scripts hasta el desarrollo de aplicaciones a gran escala. Algunos de ellos se encuentran diseñados para plataformas específicas (Murphy,2019). Estos pueden ser Visual Basic (aplicable para Windows), XCode (aplicable para macOS y iOS) y Jupyter Notebook.

En la práctica, es imprescindible contar con entornos que agilicen la gestión de los lenguajes de programación más utilizados, ya que estos ofrecen extensas bibliotecas de código que enriquecen las capacidades del software, más allá de las proporcionadas por defecto con el instalador. Estas bibliotecas, compuestas por librerías, consisten en colecciones de código compartidas por individuos u organizaciones para uso libre, permitiendo a los usuarios explorar y profundizar el lenguaje en diversas ramas informáticas de interés. Aquellas organizaciones encargadas de mantener las bibliotecas más reconocidas a menudo generan rentabilidad, ya que se vuelven indispensables para el funcionamiento de numerosas aplicaciones cotidianas. Su utilización implica la instalación a través de internet o mediante un instalador (Wardhan, 2021). Una de las librerías más usadas en el entorno de programación es Selenium (Selenium, s.f.). Esta es una herramienta de código abierto que se utiliza para automatizar navegadores web e incluye tres de los lenguajes de programación más conocidos: Python, R y Javascript[4].

[4] Es importante mencionar que también existe una librería denominada ”EvilSelenium”, la cual se utiliza para robar credenciales almacenadas (a través de autocompletar), robar cookies, tomar capturas de pantallas de sitio web, extraer y filtrar

De lo señalado, tanto los lenguajes de programación como los navegadores desempeñan un papel fundamental en la ejecución del *Web Scraping*. Este proceso se apoya en la infraestructura y la disponibilidad de información alojada en las distintas páginas web que estos proporcionan.

En esa línea, de acuerdo con el World Wide Web Consortium (W3C), una **página web** ("*web page*") es un recurso no integrado obtenido a partir de un único URI mediante HTTP (mas no limitándose a este) y cualquier otro recurso utilizado en la representación o destinado a ser representado junto con él por un agente de usuario. Esta página web puede incluir contenido multimedia, componentes interactivos y aplicaciones web (W3C, 2014).

Por otro lado, una ***website*** es el conjunto coherente de una o varias páginas web (*web pages*) relacionadas entre sí que ofrecen un uso o funcionalidad común. Esta *website* cuenta con un *website owner*, es decir, un encargado.

En esa línea, un ***usuario de internet*** es aquel que puede leer periódicos, revistas y libros, realizar operaciones bancarias o comprar productos "en línea", o comunicarse con su familia en el exterior, o conocer personas y entablar amistades (Bradley, 2000), a través del correo electrónico y de la mensajería instantánea. Esto lo aplica a todos los propósitos o intereses en su vida; desde aspectos culturales, de vicio, ocio y/o afectivos. Por ello, el hombre común, es en esta época por lo general un usuario de Internet.

Es evidente la distinta y diversa presencia de actores clave para la continuidad del Internet y, con ello, el desarrollo de millones de páginas webs en conjunto con sus navegadores. Esto sin contar a los abundantes y particulares lenguajes de programación que, sin su existencia, tampoco sería posible la navegación del Internet y, a su vez, el *Web Scraping*.

1.2 ¿Qué es el *Web Scraping*?

En medio de un mundo digitalizado, cambiante y de redes sociales, existe una inagotable fuente de información disponible en la web. Esto es, en parte, porque muchos usuarios permiten el

documentos de un ordenador, acceder a mensajes y/o correos de Whatsapp/Gmail/Office. Para mayor información visitar la siguiente página web: https://github.com/mrd0x/EvilSelenium.

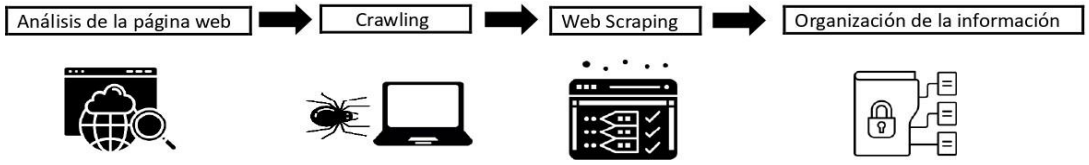
acceso a su información producida por los contenidos que generan. No obstante, es sumamente dificultoso acceder directamente, descargar, extraer y/o conservar tales contenidos de forma estructurada, lo cual a su vez hace engorroso el análisis posterior que se pueda realizar (Diouf et al., 2019).

A pesar de que tales acciones puedan realizarse manualmente, ello resulta altamente ineficiente dado el volumen y velocidad que se producen los datos. Por tanto, aparece como solucionador de este problema el *Web Scraping.*

El *Web Scraping*, o también llamado "*Web Harvesting*" (Alarcón-Urbistondo, Rojas-de-Gracia, & Casado-Molina, 2023), "*Web Data Extractio*n" (Ferrara, De Meo, Fiumara, & Baumgartner, 2014) y "*Web Data Mining*" (Singh y Singh, 2010), es la construcción de un agente (*web crawler* o *spider*) para descargar, analizar y organizar datos de la web de forma automatizada (Vanden Broucke y Baesens, 2018, p. 3–23). De acuerdo con Camargo-Henríquez y Núñez-Bernal (2022), esta técnica se utiliza para **extraer información de sitios web** de forma automatizada mediante el análisis y manipulación de la estructura HTML, de forma tal que se rastrea, recupera y estructura información incrustada en etiquetas HTML de las páginas web. Es decir, se puede definir al *Web Scraping* como el proceso de extracción de datos de páginas web que coadyuva a la generación de conjuntos de datos. Existe una multitud de posibilidades para la extracción de la información y, en muchas ocasiones, no es una tarea fácil, considerando que muchas páginas web impiden la visualización de determinada información y/o complican tal visualización para que los *scrapers* no puedan ejecutar el *Web Scraping* de manera tan sencilla.

La gran mayoría de los procesos de *Web Scraping* se pueden dividir en cuatro fases. El principal motivo de la separación de tales procesos sucede por la posibilidad de su ejecución en diferentes momentos. Generalmente, este proceso tiene una serie de fases bajo supervisión humana -para gestionar las fases- (Camargo-Henríquez; Núñez-Bernal, 2022):

Figura 5.
Flujo del *Web Scraping*



Fuente: Elaboración propia

- Fase 1: en esta fase es necesario revisar la estructura subyacente al contenido ofrecido para comprender la forma y presentación de los datos, los cuales usualmente son entregados en lenguaje de marcado HTML (lenguaje de etiquetas usualmente utilizados para navegadores) o XML (lenguaje de etiquetas más personalizado) y lenguaje de estilos CSS (relacionado al estilo de la estructura de los bloques de declaración). Cabe resaltar que los lenguajes de marcado sirven como estructura o centro para conectarse a los lenguajes de programación y de estilos. Para una mejor comprensión y desarrollo, se requiere del conocimiento de la arquitectura DOM[5] para poder identificar qué bloques y cómo se quiere extraer la información (si es por la etiqueta, el ID, la clase, los atributos, entre otros).

- Fase 2: Para rastrear las direcciones de Internet que contienen los datos que se buscan obtener se hace la carga de la *website* en una dirección inicial y se aplican reglas para identificar el resto de los hipervínculos deseados. Lo anterior se denomina *crawling* (Jarmul & Lawson, 2017). Este proceso de automatización de la extracción de la información utiliza lenguajes de programación, como Python o Java Script, que contengan librerías que faciliten la navegación de las páginas. Este *web crawler* o *spider* visita páginas y sitios web cargando, así como conservando, los elementos deseados del código HTML. Este proceso puede ejecutarse diariamente para comprobar si se tiene nueva información útil, actualizar la información general, entre otros.

- Fase 3: Es la fase principal y se aprovecha de la información obtenida en la Fase 2. Implica la creación de reglas específicas que permitan la extracción de información ordenada y sin errores de cada uno de los elementos obtenidos de la Fase 2 (de los

[5] El Modelo de Objetos de Documento (Document Object Model – DOM) es una interfaz de programación para documentos HTML y XML. Este permite acceder y manipular el contenido, estructura y estilo de una página web. Recuperado de: https://www.w3.org/DOM/faq.html#webpages.

lenguajes de programación obtenidos) que se conoce como ***Web Scraping***. Este proceso inicia con la ejecución de los *drivers* del navegador que permitirá navegar en cada uno de los elementos extraídos de la Fase 2. Dentro de cada iteración se aplicarán reglas –como la identificación de ID, patrones de estructura DOM - para identificar y extraer los datos deseados. A la par, se incluye también un proceso de contingencia ante la falta de información o información errada, lo cual implica la regularización de los bloques de datos para evitar errores en el código y así realizar la lectura correcta de los datos.

- Fase 4: Es la fase final que implica limpiar, procesar y estructurar la información para su posterior almacenamiento y análisis. Teniendo en cuenta el volumen de datos, su complejidad y estructura, es vital realizar actividades para que, luego de estructurarlos, se puede transferir a un formato de almacenamiento estándar (CSV, JSON o TXT), pudiendo utilizar lenguajes de programación (por ejemplo, R, Java, Python) para su ordenamiento y exportación del conjunto de datos.

El desarrollo de ambas herramientas –*crawling* y *Web Scraping*- sirven para recuperar conjuntos de datos tanto en la web como en la nube o en los macrodatos cuando el autor no autoriza la disponibilidad de la información (Khan et al., 2020). La diferencia es que las herramientas de *crawling* visitan páginas web para construir un índice de hipervínculos que resultan útiles posteriormente para brindar resultados de búsqueda rápidos y relevantes; mientras que el *Web Scraping* interactúa con el navegador y la aplicación web para extraer y convertir los datos encontrados en la web en un formato estructurado para su uso posterior.

A continuación, presentamos algunos inconvenientes que se pueden presentar al momento de hacer la navegación y extracción de la información mediante el *Web Scraping*:

- En la mayoría de los *websites*, la información estructurada proviene de consultas a bases de datos que nuestro navegador realiza por medio de scripts cliente/servidor. Es decir, la información no se encuentra disponible dentro de los archivos HTML que se encuentran colgados en la internet. Esto es porque las páginas se encuentran diseñadas para ser procesadas por navegadores web del cliente, consultando solo la información que necesita el usuario. Por lo tanto, si se desea obtener la mayor cantidad de información de una página, se deberá recorrer sus dominios (o simularlo).

- Otro problema es la distribución de los datos. Cada *website* posee una diferente distribución de sus elementos y no suelen ser similares a las otras *website* disponibles. Por ello, generar un mismo código que sea replicable en diferentes *website* implica un esfuerzo bastante mayor al de generar un código funcional para una página en específico.
- Un factor importante, pero menos frecuente, es que, dentro de algunas páginas de listados, los artículos aparecen de manera individual o como parte de un grupo – como parte de una oferta-. Ello genera una duplicidad de datos, salvo que el algoritmo pueda identificarlo y resolverlo.
- El valor también puede estar en diferentes idiomas con una escritura diferente.
- Otro factor es que los valores pueden contener scripts inesperados (por ejemplo, Jvascript, HTML, XML) introducidos por el usuario que, si no se tratan, pueden interrumpir la extracción de la información.
- También es necesario tener restricciones en cuanto a la cantidad y velocidad de consultas que se realizan al navegar por un *website*. Si con el apoyo de un *web crawler* las consultas por segundo son bastante elevadas por un determinado tiempo, el servidor puede detectar el comportamiento anormal y bloquear temporal o permanente al IP desde donde se esté realizando el *Web Scraping o* ejecutar acciones legales.
- Por último, es necesario actualizar el conjunto de los datos frecuentemente, pues los desarrolladores de las *website* pueden cambiar la estructura del código (Khder, 2021) (por ejemplo, el autor puede añadir, eliminar y/o modificar la estructura de la página que puede conllevar a modificar las reglas del *Web Scraping*). En concreto, los *website* se adaptan continuamente, por lo que los parámetros de *Web Scraping* utilizados para extraer los datos originales podría ser obsoleto conforme pase el tiempo.

Las situaciones descritas previamente constituyen eventualidad a considerar dentro de un proceso de *Web Scraping*. Por otro lado, se pueden distinguir distintos modelos de implementación de *Web Scraping* para cada escenario:

A. Mimetismo: Este modelo utiliza reglas predeterminadas desde una plantilla que recrea o imita los selectores DOM del contenido original que se va a analizar (Gupta et al., 2003). No obstante, su efectividad se ve reducida al utilizarlo en diferentes tipos de *website*, por lo que se necesita un criterio en la programación a desarrollar para poder

discriminar los datos necesarios. Import.io (Import.IO, 2022) o Mozenda (Baskaran and Ramanujam, 2018, p. 19) son herramientas que utilizan este modelo.

B. Aprendizaje automático: Este modelo busca utilizar una gran cantidad de la muestra de contenidos analizados de manera manual a través de un algoritmo, que mediante el aprendizaje automático analiza, define y deduce dónde se encuentra el bloque de texto principal comparándolo con los demás bloques de texto (Vargiu y Urru, 2013). Por tanto, mientras mayor es la muestra, más preciso el algoritmo. Sin embargo, este modelo constituye un mayor esfuerzo computacional y de los programadores para su ejecución al ser masivo.

1.3 Conceptos claves para la ejecución del *Web Scraping*

Con el propósito de entender a profundidad las implicancias del *Web Scraping*, este apartado reúne una introducción a todas las tecnologías que influyen en la realización de un proyecto de *Web Scraping*.

1.3.1. HTML (lenguaje DOM)

Otro elemento fundamental relacionado con el W*eb Scraping* es HTML (Lenguaje de Marcas de Hipertexto), el cual es un lenguaje informático sobre el que se desarrolla la estructura de la mayoría de las *website* y aplicaciones en línea.

Para crear cualquier documento HTML se necesitan de las etiquetas, las cuales indican qué estructura va a contar la *website*. El lenguaje DOM indica la secuencia de etiquetas que se debe seguir para llegar a una ubicación de la página en específico, lo que permite al *crawler* identificar la información a extraer. Las etiquetas más comunes que se pueden emplear en un código de *Web Scraping* son las siguientes:

Tabla 1.
Principales etiquetas HTML

Etiqueta	Descripción
<body>	Elemento tiene dentro todo el contenido visible de la *website*.
<h1> <h2>...<h7>	También llamado "*Heading*". Se utiliza para crear títulos, subtítulos, entre otros equivalentes. Existen desde el h1 hasta el h7.
<p>	"Paragraph" elemento define un párrafo. Se utiliza para escribir textos largos y simples.
<table>	Estructura de tabla dentro de la *website*.

	Dividir contenido
<a>	Hipervínculos o enlaces
	Formato de texto en negrita
 	Saltos de línea
<img>	Añadir imágenes en un *website*
	Sirve para ordenar listas
	Sirve para listas desordenadas
	Para elementos en una lista

Fuente: Elaboración Propia

1.3.2. Jupyter Notebook

Resulta imprescindible profundizar en Jupyter Notebook, pues es el entorno de desarrollo del *Web Scraping*. Este versátil entorno no se limita únicamente a la ejecución de Python, ya que también es compatible con R y puede programarse para ejecutar otros lenguajes, como Julia. Se le reconoce en gran medida por su inclusión en las aplicaciones de Anaconda, una distribución integral para lenguajes de computación científica. Además, su popularidad se ve respaldada por su capacidad para ejecutar código en bloques, lo que posibilita la ejecución selectiva de partes del código en lugar de ejecutar el conjunto completo en cada iteración. Esta característica resulta especialmente conveniente en la programación con datos, puesto que realizar operaciones repetitivas sobre conjuntos de datos extensos puede consumir recursos significativos. Para instalarlo, deberán seguir las indicaciones de su página web oficial.

1.3.3. Librería Selenium

Selenium es una librería de código abierto diseñada para la automatización e interacción de navegadores web. Su principal propósito es realizar simulación de la interacción humana con un navegador, lo que permite interactuar con todas sus partes.

Selenium, a su vez, soporta varios lenguajes de programación, como Python, R y JavaScript, lo que lo hace atractivo para los desarrolladores. Además de la automatización de pruebas, Selenium es utilizado comúnmente en tareas como **la extracción de datos web** (incluyendo el *Web Scraping*), la realización de pruebas de rendimiento y la automatización de procesos repetitivos en entornos web.

1.3.4. Controlador de navegador web

Los controladores de navegadores web son herramientas que proporcionan parte de los documentos, código e interfaz de navegadores web para el uso en pruebas y desarrollo. Un controlador de navegador web muy usado por desarrolladores es *ChromeDriver*. Este último permite la simulación de la interfaz en específico de Google Chrome que se utiliza en el proyecto elaborado en el acápite 1.4.

Su funcionamiento es ejecutable desde el lenguaje de programación deseado. De esta forma, *Chrome driver* permite la automatización de acciones y la manipulación de elementos en páginas web.

1.4 Simulación del *Web Scraping*

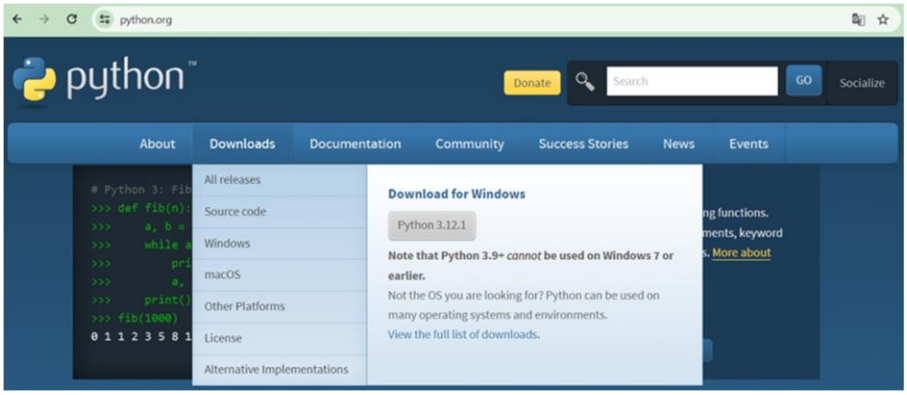
A fin de aterrizar y materializar el proceso del *Web Scraping*, se procederá a explicar detenidamente cómo se puede ejecutar.

> **Objetivo**: Obtener un conjunto de *dataset* estructurado de los Proyectos de Ley del Congreso del 2021 (Congreso de la República, s.f)

Pasos previos:

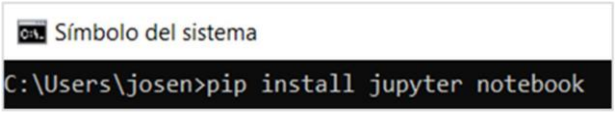
1. Entramos a instalar Python: https://www.python.org/.

Figura 6.
Página web Python



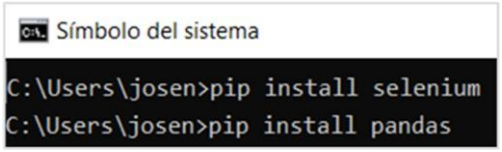
2. Entramos al CMD (Símbolo del sistema) e introducimos el siguiente comando: "*pip install jupyter notebook*". Esto descargará el entorno de desarrollo Jupyter, en donde estaremos programando nuestro código.

Figura 7.
CMD (Símbolo del sistema)



3. Dentro del CMD ejecutar también los comandos: "*pip install Selenium*" y "*pip install Pandas*", que son las librerías externas que utilizaremos para nuestro programa.

Figura 8.
Inclusión de librerías externas en el CMD (Símbolo del sistema)



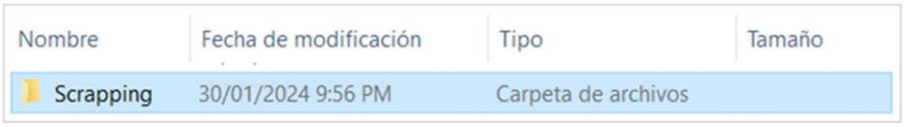
4. Descargar Google *Chrome Driver*[6] según el tipo de versión de Chrome. Esta herramienta hará posible las pruebas dentro de un navegador Google Chrome.

Figura 9.
Página web del Chrome Driver

[6] Acceder mediante el siguiente link: https://chromedriver.chromium.org/downloads.

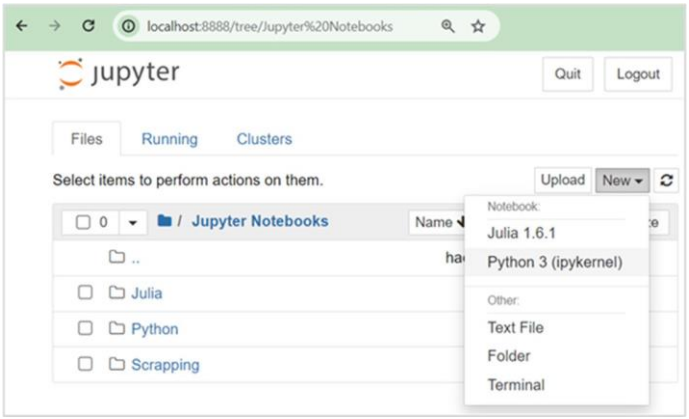
5. Crear una carpeta donde se ejecutará el código. Tener en cuenta que todos los archivos que se generen se guardarán en tal carpeta.

Figura 10.
Carpeta con la ejecución del código



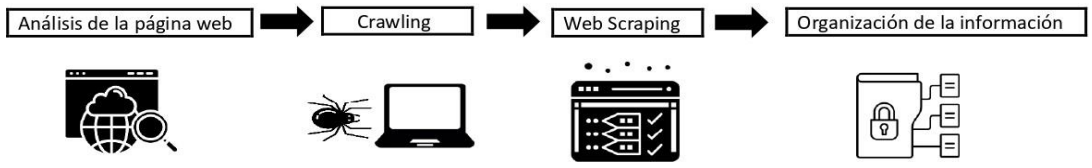
6. Crear un nuevo notebook de Python en Jupyter.

Figura 11.
Jupyter



Recordando los pasos para ejecutar el *Web Scraping*, detallamos el paso a paso de cada uno a fin de completar el objetivo:

Figura 5.
Flujo del *Web Scraping*



Fuente: Elaboración propia

- Bloque 0: Se instalan las librerías a utilizar (solo se ejecuta una vez por computadora)

```
#instalamos las librerias
#pip install selenium
#pip install pandas
```

- Bloque 1: Se importan las librerías a utilizar

```
from selenium import webdriver
from selenium.webdriver.common.by import By
import time
import pandas as pd
```

Paso 1: Análisis de la *website*

Se identifica la *website* inicial incluyendo la escala y la estructura del sitio web de destino como la ubicación de los enlaces que se van a obtener y la estructura de la información deseada por proyecto.

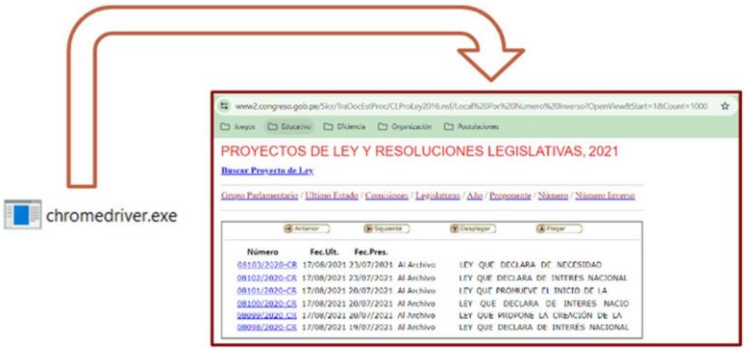
Por un lado, la escala del sitio web o tamaño afectará la forma en cómo se va a realizar el *crawling* o rastreo de información. Por otro, la estructura del sitio web que se desea *scrapear* sirve para (i) visualizar sus robots.txt y conocer las sugerencias de la estructura del sitio web, indicar a los *crawlers* o rastreadores de los buscadores a qué URLs del sitio web pueden acceder sin ser bloqueado, (R. Larson, 2015); (ii) mapas del sitio, que son archivos que sirven para ayudar a los motores de búsqueda a buscar, rastrear e indexar el contenido de un sitio web; y (iii) las herramientas externas disponibles para visualizar los detalles, como Google y el protocolo WHOIS.

Figura 12.
Análisis de la *website*

Seguidamente, se ejecuta el controlador web "*Chrome driver"* e iniciamos en la *website* que recopila los Proyectos de Ley del Congreso del 2021.

```
driver=webdriver.Chrome()
driver.maximize_window()
link="https://www2.congreso.gob.pe/Sicr/TraDocEstProc/CLProLey2016.nsf/Local%20Por%20Numero%20Inverso?OpenView&Start=1&Count=1000"
driver.get(link) #Entramos a la página inicial
```

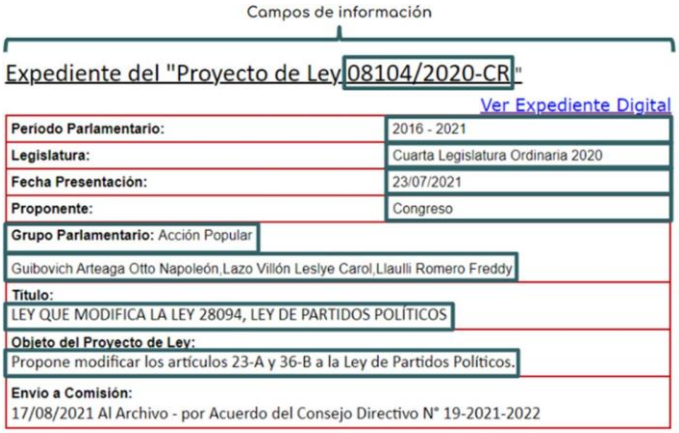
Figura 13.
Ejecución del Chrome Driver



Luego, se identifica en la estructura DOM los campos de información de cada Proyecto de ley.

```
#Identificamos en la estructura DOM los campos de información de cada proyecto de ley
periodo='//table[not(@id) or not(@class)]/tbody/tr[2]/td/table/tbody/tr[1]/td[2]'
legislatura='//table[not(@id) or not(@class)]/tbody/tr[2]/td/table/tbody/tr[2]/td[2]'
fecha_P='//table[not(@id) or not(@class)]/tbody/tr[2]/td/table/tbody/tr[3]/td[2]'
proponente='//table[not(@id) or not(@class)]/tbody/tr[2]/td/table/tbody/tr[4]/td[2]'
titulo = '//table[not(@id) or not(@class)]/tbody/tr[2]/td/table/tbody/tr[6]/td[1]/font'
sumilla = '//table[not(@id) or not(@class)]/tbody/tr[2]/td/table/tbody/tr[7]/td[1]/font'
código='//table[not(@id) or not(@class)]/tbody/tr'
autores = '//table[not(@id) or not(@class)]/tbody/tr[2]/td/table/tbody/tr[5]/td[1]'
#Los almacenamos en una lista
xpaths = [código, periodo, legislatura, fecha_P, proponente, titulo, sumilla, autores]
```

Figura 14.
Identificación de la estructura DOM



Paso 2: *Crawling*

Se realiza el *crawling* extrayendo los enlaces de los proyectos de ley y recopilando los links en una lista.

```
i='//table/tbody/tr/td/font/a["href"]' #Estructura DOM de los links
webElements=driver.find_elements(by=By.XPATH, value=i) #Extraemos los elementos web
links=[e.get_attribute('href') for e in webElements] #Almacenamos los links en una lista
```

Figura 15.
Recopilación de links para el *Crawling*

Paso 3: *Web Scraping*

Se crea la base de datos en donde se almacenará la información:

```
data = {'código': [],'periodo': [],'legislatura': [],'fecha_P': [],
        'proponente': [],'titulo': [],'sumilla': [],'autores': []}
```

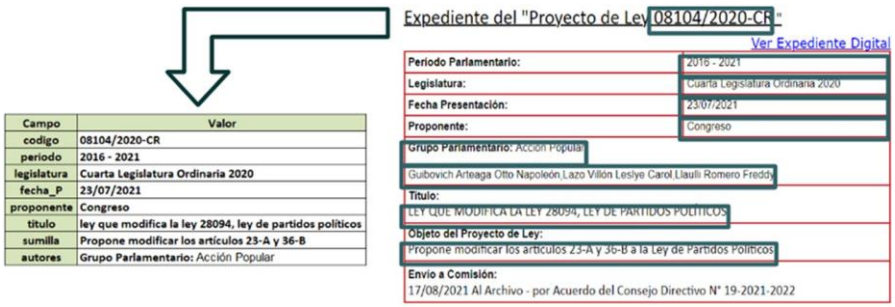
Figura 16.
Base de datos para almacenar la información

	codigo	periodo	legislatura	fecha_P	proponente	titulo	sumilla	autores
0								
1								
2								
3								

Luego, se recorre cada enlace de los Proyectos de ley y extraemos los campos de información:

```
#Recorremos cada uno de los links
for link in links:
    #Esperamos 1 segundo por proyecto de ley para que el servidor de la página del congreso no bloquee nuestro IP
    driver.implicitly_wait(1000)
    time.sleep(1)
    driver.get(link)
    #Extraemos todos los campos necesarios
    for path,i in zip(xpaths,range(len(xpaths))):
        x=driver.find_element(by=By.XPATH, value=path).text
        data[list(data.keys())[i]]=data[list(data.keys())[i]]+[x]
```

Figura 17.
Identificación de la extracción de la información



Paso 4: Organización de la información

Se convierte nuestro conjunto de información en un *dataframe*, entendido como una estructura de datos en el que se puede guardar datos de distintos tipos.

```
#Convertimos nuestro dataset en un dataframe
df = pd.DataFrame(data)
```

Posteriormente, se recorre cada enlace de Proyectos de ley y extraemos los campos de información:

```
Limpiamos la información (columna codigo)
df['codigo']=df['codigo'].str.replace('Expediente del "Proyecto de Ley ','").replace(' "',")
```

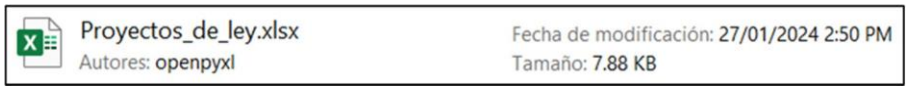
Figura 18.
Extracción de los campos de información

codigo
Expediente del "ProyectodeLey08104/2020-CR"
Expediente del "ProyectodeLey08103/2020-CR"
Expediente del "ProyectodeLey08102/2020-CR"
Expediente del "ProyectodeLey08101/2020-CR"
Expediente del "ProyectodeLey08100/2020-CR"
Expediente del "ProyectodeLey08099/2020-CR"
Expediente del "ProyectodeLey08098/2020-CR"
Expediente del "ProyectodeLey08097/2020-CR"
Expediente del "ProyectodeLey08096/2020-CR"
Expediente del "ProyectodeLey08095/2020-CR"

codigo
08104/2020-CR
08103/2020-CR
08102/2020-CR
08101/2020-CR
08100/2020-CR
08099/2020-CR
08098/2020-CR
08097/2020-CR
08096/2020-CR
08095/2020-CR

Finalmente, extraemos la base de datos a un archivo Excel:

Figura 19.
Extracción de la información



1.5 ¿Es el *machine learning* capaz de elaborar códigos aplicables al *Web Scraping*?

Con el propósito de responder adecuadamente la pregunta previa, es necesario entender a *prima facie* qué es inteligencia artificial (también denominado por sus siglas "**IA**"). El término fue acuñado por John MacCarthy en 1955 y definía a la IA como "*la ciencia y la ingeniería de hacer máquinas inteligentes*". Asimismo, Alan Turing diseñó una especie de prueba para determinar cuándo nos encontrábamos frente a una solución de IA (Russell y Norvig, 2004).

Sobre el particular, a la fecha no existe un consenso sobre su definición. Por un lado, Delgado sostiene que la inteligencia artificial es la ciencia que busca crear sistemas y programas capaces de imitar la inteligencia humana, permitiendo a las máquinas aprender, razonar y tomar decisiones autónomas (Delgado, s.f.). Así, la IA puede entenderse como la elaboración de sistemas informáticos y algoritmos que realicen tareas que, en su mayoría, necesitan de inteligencia humana (Pedrero, 2023). Por ejemplo: la traducción de idiomas o la toma de decisiones complicadas.

De acuerdo a Bellman, (1978, citado por Russell y Norvig, 2004), la IA es la automatización de actividades que vinculamos con procesos de pensamiento humano, actividades como la toma de decisión, resolución de problemas y aprendizaje.

La Organización para la Cooperación y el Desarrollo Económicos (en adelante, "OCDE"), conforme a su última modificación (OCDE, 2023), define a los sistemas de IA como aquel sistema basado en máquinas que, por objetivos explícitos o implícitos, infiere, a partir de la entrada que recibe, cómo generar salidas, como pronósticos, contenidos, consejos o decisiones que pueden influir en escenarios físicos o virtuales. Como parte de los principios que deben ser tomados en consideración para el manejo de la IA, se sostiene que los actores que manejen sistemas de IA deben respetar los principios básicos del derecho, derechos humanos y valores democráticos, en los que se incluyen la privacidad y los datos personales.

En esa misma línea, una publicación reciente de la Fundación Telefónica (2023), la UNESCO define a la inteligencia artificial como las máquinas capaces de imitar determinadas funcionalidades de la inteligencia humana, incluidas características como la percepción, el aprendizaje, el razonamiento, la resolución de problemas. Además, Fernández Naranjo (2023) discute el impacto de la inteligencia artificial, citando a Benitez López (2022) quien destaca que la IA necesita estímulos que son los datos que serán manejados y procesados por algoritmos, los cuales son las instrucciones que reciben las máquinas. Estas funcionan gracias al *machine learning* o aprendizaje automático.

Asimismo, la Unión Europea ("**UE**") aprobó, recientemente, la Ley de Inteligencia Artificial (Unión Europea, 2021), en el cual, si bien aún no ha entrado en vigor como tal y se espera que próximamente, dicha publicación y consenso de los Estados miembros de la UE ha representado un avance histórico para regulación de dicha tecnología. En ella se define a la inteligencia artificial como aquel software desarrollado con una o varias de las técnicas[7] para un conjunto determinado de objetivos definidos por el ser humano, y que, a su vez, puede

[7] Las técnicas identificadas por la norma son las siguientes:
(i) Enfoques de aprendizaje automático (*machine learning*), incluido el aprendizaje supervisado, no supervisado y de refuerzo, utilizando una amplia variedad de métodos, incluido el aprendizaje profundo (*deep learning*);
(ii) Enfoques basados en la lógica y el conocimiento, incluida la representación del conocimiento, la programación inductiva (lógica), las bases de conocimiento, los motores de inferencia y deducción, el razonamiento (simbólico) y los sistemas expertos;
(iii) Enfoques estadísticos, estimación bayesiana, métodos de búsqueda y optimización.

generar resultados como contenidos, pronósticos, consejos o decisiones que influyen en los entornos con los que interactúan.

Finalmente, el Estado peruano define a la IA como aquella tecnología emergente que tiene el potencial de mejorar el bienestar de las personas, contribuir a una actividad económica global sostenible positiva, aumentar la innovación y la productividad, y ayudar a responder a los desafíos globales clave (Ley No. 31814, 2023, Art. 3). En adición a ello, según la Ley No. 31814 (2023, Título Preliminar- Artículo Único) se resalta como principio para el desarrollo y uso de la IA que, esta última, (i) no debe trasgredir la privacidad de las personas, y (ii) debe actuar de manera segura para lograr un impacto positivo y bienestar en los ciudadanos. De acuerdo a la Estrategia Nacional de Inteligencia Artificial (en adelante, "ENIA"), se tiene previsto la implementación del Centro Nacional de Innovación e Inteligencia Artificial (ENIA, 2021), que tiene, como parte de sus funciones, la realización y publicación de investigaciones a la IA, promover alianzas con instituciones nacionales y extranjeras, generar recomendaciones en coordinación con el sector público, privado y académico del país, y patrocinar eventos relacionados a la IA.

El Proyecto de Reglamento de la Ley de IA[8] indica que, durante todas las etapas del ciclo de vida de la IA, se debe respetar, promover y proteger la dignidad, autonomía y los derechos fundamentales, incluyendo la protección de los datos personales, existiendo la responsabilidad ética, civil, administrativa y penal por el uso y desarrollo de los sistemas basados en IA. De ello, se sostiene que el implementador de un sistema basado en IA debe informar al ciudadano o consumidor, de forma clara y sencilla, el uso del sistema basado en IA con el que interactúe o con el servicio o producto que utiliza a la IA como uno de sus componentes, lo cual también incluye los parámetros utilizados para llegar a una decisión o resultado obtenido al utilizar la IA para la generación de predicciones, contenido, recomendaciones o la toma de decisiones.

La IA cuenta con distintas definiciones a la fecha; sin embargo, todas ellas coinciden en que se trata de una serie de técnicas y/o mecanismos que tratan de simular el comportamiento humano teniendo como resultado final un resultado: reportes, estadísticas, datos, productos, códigos,

[8] Se tuvo acceso al presente documento mediante la Solicitud de Acceso a la Información Pública realizada a la Secretaria de Gobierno y Transformación Digital (SEGDI) con fecha 29 de enero de 2024. A la fecha de redacción de la presente tesis, la SEGDI tiene prevista la publicación del Reglamento de la Ley de IA, cuyo contenido también contemplará los riesgos entre los datos personales y el uso de la inteligencia artificial, en especial por casos de *deepfakes* y suplantación de identidad. Recuperado de: https://andina.pe/agencia/noticia-apec-2024-pondra-foco-transformacion-digital-inteligencia-artificial-y-ciberseguridad-971021.aspx.

entre otros. Siendo ello así, una de las técnicas más particulares de la IA es la de enfoques de aprendizaje automático, también denominado "*machine learning*". De acuerdo con Bellman, (1978, citado por Russell y Norvig, 2004), este es subcampo de la IA que automatiza la construcción de modelos analíticos utilizando redes neuronales, estadísticas, investigación de operaciones y física para hallar los conocimientos en los datos sin programar explícitamente donde buscar o qué concluir.

Así, se entiende como *aprendizaje automático o machine learning* a todo aquel conjunto de modelos de IA que aprenden en base a datos (de entrenamiento) para poder predecir resultados o tomar decisiones sin ser explícitamente programados para ello (ENIA, 2021). En otras palabras, se define como la capacidad de las máquinas de aprender de los datos para resolver una tarea sin ser programadas para ese encargo específicamente (Muñoz, 2022).

En ese sentido, para cualquier modelo de *machine learning* habitualmente se dividen los datos disponibles en tres conjuntos: (i) datos de entrenamiento (*training*), (ii) datos de validación (*validation*) y (iii) datos de prueba (*test*). Los datos de entrenamiento son los que se usan para que algoritmo de aprendizaje pueda obtener los parámetros del modelo. Podemos realizar ajustes de los hiperparámetros con los datos de validación hasta que obtengamos unos resultados de validación que consideremos correctos. Finalmente, los datos de prueba se utilizan para la evaluación final del modelo dentro de todo el proceso, cuando se considera que el modelo está listo y ya no se modificarán ninguno de sus hiperparámetros (Torres, 2018).Este método es netamente basado en la utilización de algoritmos que, en base a sus propias características y a la gran cantidad de datos para aprender pueden ser objeto de amenazas jurídico-legales como: comprometer la privacidad durante las operaciones de datos, corrupción de índices de datos, perfilado de usuarios finales, falta de cumplimiento de la protección de datos de terceros, dependencia de un proveedor, falta de datos políticas gubernamentales o divulgación de información personal (Muñoz, 2022).

Dicho ello, respondiendo el acápite que nos llevó al análisis en cuestión, el *machine learning* o aprendizaje automático sí puede generar códigos aplicables al *Web Scraping*. Esto es porque basta con solo enviarle una instrucción detallada al software para que pueda brindar el código específico. Dicho código debe ser representado de forma tal que pueda ser entendido por el modelo de *machine learning*, lo que se puede realizar convirtiendo el código en una secuencia de tokens o estructura de datos para el procesamiento.

Luego, este modelo se entrena utilizando del conjunto de datos etiquetado y, conforme se vaya dando los entrenamientos, el modelo aprende a asociar las características de las páginas web con el código de extracción aplicable.

Uno de los ejemplos de ejecución de *Web Scraping* a través de mecanismos de *machine learning* es con la solución de ChatGPT[9]. En sí, ChatGPT es un modelo de lenguaje desarrollado por OpeanAI de uso libre que está diseñado para generar respuestas textuales similares a los humanos a partir de la información que reciben.

De ese modo, ChatGPT puede generar códigos Python para realizar *Web Scraping* a través de su motor Code Interpreter. Por ejemplo, si le pedimos "Scrapear XXX para títulos de libros usando Python y BeautifulSoup" entonces aparecerá el siguiente código:

```
# Import necessary libraries
import requests
from bs4 import BeautifulSoup
# Make a request to the website
r = requests.get('XXX')
# Create a BeautifulSoup object soup = BeautifulSoup(r.text, 'html.parser')
# Find book titles titles = soup.find_all('h3') for title in titles: print(title.get_text()
```

Como se puede evidenciar, este es uno de los ejemplos de lo que puede hacer el *machine learning* en relación con el *Web Scraping*. En base a lo revisado, consideramos firmemente que no será la primera ni la última solución que permita realizar esta técnica a través del *machine learning*, siendo que, cada vez más, existe una mayor aceptación en soluciones de IA. Por ejemplo, de acuerdo a un reciente reporte de Stanford University (Stanford University, 2023), el Perú se encuentra dentro de los cinco países que tienen mayor aceptación a la IA, por lo cual cada vez más se considera como beneficio a todo producto y/o servicio que contenga IA.

CAPÍTULO II. CONCEPTO DE LOS DATOS PERSONALES Y LA REGULACIÓN PERUANA

2.1 Naturaleza del dato

[9] Para mayor información, accede al siguiente link: https://chat.openai.com/.

Desde su rasgo etimológico, la palabra dato proviene del latín “datum” que significa "*(cosa) dada*" (Harper, s.f.). Con la evolución del tiempo, en 1946 se consideraba dato a toda aquella información transmisible y almacenable mediante la cual se realizan operaciones informáticas (Harper, s.f.). Tan así se ha desarrollado el campo de los datos que, a la fecha, existe una rama de la ciencia que estudia los datos *per se*, siendo la ciencia de datos (o también denominado ”*data science“*) el conjunto de herramientas que ayuda a los gestores a basar sus decisiones en pruebas (Prevos, 2019).

De acuerdo con el *Cambridge English Dictionary* (s.f.), se entiende por ”dato” como aquella información, especialmente hechos o cifras, recopilada para ser examinada, considerada y utilizada como ayuda para la toma de decisiones. Asimismo, se podría entender con dato a todo aquello que define distintas cosas, pero a la voz genera incertidumbre: un todo o nada. De acuerdo a Puccinelli (1997), el vocablo dato hace referencia a un elemento circunscripto y aislado que no tendrá carácter de información, pues para ello se requeriría la interconexión de dichos datos que vinculados se transformen en una referencia determinada.

El Estado peruano define al dato como aquella representación dimensionada y descifrable que abarca: hechos, información o conceptos. Todos estos expresados en cualquier forma aplicable a su almacenamiento, procesamiento, interpretación y comunicación (Secretaría de Gobierno Digital, 2018) [10]. Aunado a ello, incluso realiza una subclasificación señalando que existe el término de *datos abiertos por defecto* entendido por quienes se encuentran disponibles de manera inmediata, sin comprometer el derecho de protección de datos de los ciudadanos y, ante la duda, deja a la Autoridad de Transparencia para que pueda definirlo (Decreto Legislativo No. 1412, 2018).

Siendo ello así, existe una serie de definiciones que se ha otorgado al término dato; sin embargo, nos inclinamos a la posición de Kitchin (2014) que determina como dato a la materia prima generada al clasificar el mundo en categorías, medidas y diversas formas de representación, tales como números, caracteres, símbolos, imágenes sonidos, que son los componentes elementales para la creación de información y conocimiento.

[10] También se encuentra definido en el Decreto Legislativo No. 1412 (Art. 23.1), Decreto Legislativo que aprueba la Ley de Gobierno Digital.

En esa línea, los datos existen inclusive antes de la argumentación o interpretación que los convierte en hechos, pruebas e información (Rosenberg, 2013). Asimismo, siguiendo con la naturaleza del dato, estos pueden ser catalogados de distintas formas (Kitchin, 2014), como las siguientes:

- De naturaleza representativa: en base a mediciones de un fenómeno como la edad, altura, peso, tensión arterial, los hábitos, la ubicación, entre otros)
- Implícitos: ausencia en lugar de una presencia
- Derivados: los datos pueden ser producidos a partir de otros, como, por ejemplo, la variación porcentual extraída de una gran base de datos.
- Registrarse y almacenarse de forma analógica o codificarse de forma digital en forma de bits (dígitos binarios).

Visto lo anterior, la data variará dependiendo de la forma (cualitativa y cuantitativa), estructura (estructurada, semiestructurada, o no estructurado), fuente (derivado, capturado, transitorio), productor (primario, secundario, terciario) y del tipo (metadata, indexado) (Kitchin, 2014).

Siguiendo con las estructuras que soportan los datos, Floridi (2008) sostiene que la independencia del soporte de los datos depende de tres tipos de neutralidad:

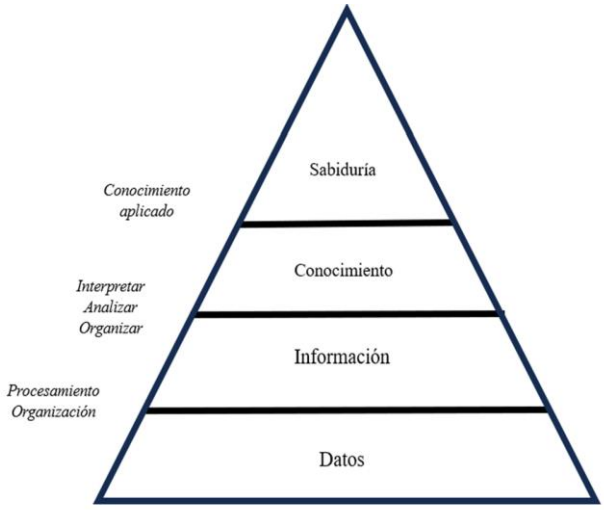
- Taxonómica: los datos son entidades relacionales definidas con respecto a otros datos específicos.
- Tipológica: los datos pueden adoptar diversas formas no mutuamente excluyentes, por ejemplo, primarios, secundarios, metadatos, operativos, derivados.
- Genéticos: los datos pueden tener una semántica independiente de su comprensión. Por ejemplo, los jeroglíficos egipcios constituyen datos, sin perjuicio de que cuando fueron descubiertos nadie podía llegar a su interpretación.

Así, los datos de buena calidad son discretos e inteligibles (cada dato es individual separado y separable, y claramente definido), agregativos (se pueden construir en conjuntos), tienen metadatos asociados (datos sobre los datos), y se pueden vincular a otros conjuntos de datos para proporcionar perspectivas que no están disponibles en la actualidad (Rosenberg, 2013).

De lo señalado, podemos concluir que los datos no existen independientemente de las ideas, técnicas, tecnologías, personas y contextos que los producen, procesan, gestionan, analizan y almacenan (Martini, 2015). De ese modo, se formula la siguiente la pregunta: ¿todo tipo de información es un dato? La Real Academia Española (s.f.) define a la información como aquel medio comunicativo o de obtención de conocimientos que habilita ampliar o precisar una materia determinada. Es particular la definición sostenida por la RAE, en tanto coincide con lo señalado por Kitchin (2014): los datos preceden a la información que, a su vez, precede al conocimiento.

Por otro lado, Weinberger (2011), citando a Adler (1986), sostiene que existiría una especie de *Pirámide del conocimiento* que explicaría los aspectos derivados de los datos, conforme se puede apreciar a continuación:

Figura 20.
Pirámide del conocimiento



Fuente: Basado en la Pirámide de Conocimiento de Adler (1986)

Cada uno de los niveles cuenta con elementos diferentes que debe ser tomado en consideración a fin de entenderlo en su totalidad (Kitchin, 2014):

- Datos: se trata de los datos como tal, en cuyo caso son datos organizados y enmarcados en un sistema de medida.

- Información: Este valor se obtiene ordenando, clasificando, enlazando o añadiendo contenido semántico a los datos mediante algún tipo de texto o visualización que informe sobre algo o indique qué hacer. Conforme lo indica la Secretaría de Gobierno y Transformación Digital (2018), la información se refiere a los datos que se han modelado en una forma significativa y útil para los seres humanos (interpretación de los datos). Algunos expertos indican (Floridi, 2010) que la información es una mercancía con características especiales al añadir valor a los datos: (i) no excluible: se comparte fácilmente y se requiere algún tipo de esfuerzo para limitarlo (ejemplo: derechos de propiedad intelectual, secreto empresarial); (ii) coste marginal cero: una vez que la información está disponible, el costo de reproducción es casi insignificante y (iii) no rival: más de una entidad puede poseer la misma información (a diferencia de los bienes materiales). Siendo así, la información es el significado que toman los datos de acuerdo a las vinculaciones que se cuentan con estos.

  De esa manera, podemos señalar que la información se trata de los datos más un significado o valor que se le puede atribuir. Por ello, se puede señalar que se trata de datos estructurados.
- Conocimiento: el conocimiento es toda información procesable y que implica aplicar la información en procesos cognitivos complejos como la síntesis, la extracción, la asociación del razonamiento y la percepción. Tiene más valor que la información, en tanto proporciona la base para comprender, explicar y extraer ideas en general, las cuales pueden ser utilizables para acciones posteriores.
- Sabiduría: es la aplicación del conocimiento de manera eficiente, clave y oportuna.

La clasificación previa nos permite diferenciar entre lo que se considera dato con información. En sí, el dato no podrá vincularse a la información si es que este no transmite conocimiento (Chanamé, 2003). Incluso el mismo Estado peruano identifica esta diferencia (Secretaría de Gobierno Digital, 2018), señalando que toda entidad pública debe reconocer que los datos e información (en formato físico y/o digital) son un activo estratégico, desde su creación hasta disposición final. Así, también reconoce como principio a los datos como un activo estratégico para diseñar políticas, crear servicios digitales y tomar de decisiones (D.U. No. 006-2020, 2020).

Siendo esto así, surge la siguiente pregunta: ¿cuándo un dato será personal? Ello será analizado y detallado en el siguiente acápite.

2.2 Datos personales: definición y alcance

Para poder comprender qué se entiende por dato personal, es necesario dirigirnos a lo que comprende tal concepto desde la doctrina autorizada. Según Santamaría Ramos (2011, pp. 512-513), se define al dato como la unidad básica de la sociedad de la información y, por tanto, permite maximizar sus beneficios a cualquier organización al momento de recolectarlos.

Aunado a lo anterior, el dato representa un elemento de la identidad de la persona que, en conjunto con otros datos, sirve para identificarla a ella y solo a ella, y sería susceptible de usarse para coartarla, es de su propiedad, en el sentido de que tendría ciertos derechos sobre su uso. Estos involucran datos respecto a relaciones de propiedad y de familia, aspectos de su personalidad, y señales de identidad de diversa índole que van emergiendo en las actividades de la vida. Todos estos datos combinados en un modelo son equivalentes a una "huella digital" porque el individuo es identificable a través de ellos (Corte Constitucional de Colombia, 1992).

Del mismo modo, Flores Dapkevicius (2004) entiende al dato personal como la información de distinta naturaleza relacionada con personas naturales o entidades determinadas o determinables. Por ejemplo: el nombre, sexo, nacionalidad, domicilio, estado civil, inscripción en una mutualista de atención médica, número de afiliados a la seguridad social, etc.

En cuanto a la regulación internacional, existen una serie de documentos que contienen propuestas a lo que se entiende como dato personal:

Tabla 2.
Documentos internacionales en materia de protección de datos personales

Documentos internacionales	Definición de dato personal
Directrices OCDE de 1980	“Datos Personales”: Toda información correspondiente a una persona identificada o identificable (sujeto de los datos) (literal b del numeral 1)
Convenio 108 del Consejo de Europa (1981)	“Datos de carácter personal”: Cualquier información que corresponda a una persona física identificada o identificable (persona concernida) (literal a del artículo 2)

Principios de Puerto Seguro, USA (2000)	"Datos Personales e información personal": Datos que correspondan a una persona identificada o identificable que se encuentre en el ámbito de la Directiva y sean recibidos desde la UE por entidades estadounidenses, cualquiera que sea la forma en que se registren".
Marco de Privacidad APEC (2005)	"Información personal": Cualquier información que corresponda a un individuo identificado o identificable (numeral 9 de Definiciones).
Directrices de la Red Iberoamericana de Protección de Datos (2007)	"Datos de carácter personal": Cualquier información que corresponda a personas naturales identificadas o identificables (numeral 1.1.)
Propuesta de Reglamento general de protección de datos del Parlamento Europeo y del Consejo (2012)	"Datos Personales": Toda información sobre un interesado (numeral 2 del artículo 4).
Directrices OCDE sobre protección de la privacidad y flujos transfronterizos de datos personales (2013)	Mantiene la misma denominación y definición de las directrices de 1980 (literal b del numeral 1).
Guía Legislativa sobre la Privacidad y la Protección de Datos Personales en las Américas (2015)	"Datos personales": Información que identifica o permite razonablemente identificar a una persona de forma directa o indirecta, por referencia a un número de identificación o a uno o más factores relacionados específicamente a su identidad física, fisiológica, mental, económica, cultural o social.
Estándares de protección de datos personales para los Estados Iberoamericanos (2017)	"Datos Personales": Cualquier información concerniente a una persona natural identificada o identificable, expresada en forma numérica, alfabética, gráfica, fotográfica, alfanumérica, acústica o de cualquier otro tipo. Se considera que una persona es identificable cuando su identidad pueda determinarse directa o indirectamente, siempre y cuando esto no requiera plazos o actividades desproporcionadas (numeral c de Definiciones).
Recomendaciones para el tratamiento de datos personales mediante servicios de computación en la nube (2021)	Mantuvo la misma definición que los Estándares de protección de datos personales para los Estados Iberoamericanos (2017).
Principios Actualizados sobre la Privacidad y la Protección de Datos Personales, con Anotaciones (2021)	"Datos Personales": La información que identifica o permite razonablemente identificar a una persona física de forma directa o indirecta, por referencia a un número de identificación, datos de localización, un identificador en

	línea pudiendo ser uno o múltiples factores vinculados a su identidad física, fisiológica, genética, mental, económica, cultural o social. Incluye información expresada de manera numérica, alfabética, gráfica, fotográfica, alfanumérica, acústica, electrónica, visual o de cualquier otro tipo. El concepto no abarca a aquella información que no identifique a una persona en particular (o no puede usarse de manera razonable para identificarla).
Reglamento del Parlamento Europeo y del Consejo sobre normas armonizadas para un acceso justo a los datos y su utilización (2023)	"Datos Personales": Definidos en el Reglamento (UE) 2016/679 (artículo 4.1).

Ahora bien, es pertinente remitirnos a las categorías de los datos personales. Primero, para la regulación peruana, se cuenta con los siguientes tipos de datos personales:

- Datos personales en sentido estricto: Constituye todo aquel dato que permite identificar a una persona natural o la hace identificable. Incluye los siguientes conceptos: nombre, número telefónico, imagen, voz, documento nacional de identidad, pasaporte, firma, domicilio, correo electrónico, entre otros. Para este tipo de datos personales generales se puede otorgar el consentimiento vía escrita y oral.

- Datos sensibles: Se configuran por las "*características físicas, morales o emocionales, hechos o circunstancias de su vida afectiva o familiar, hábitos personales que corresponden a la esfera más íntima, la información relativa a la salud física o mental u otras que afecten su intimidad*" (numeral 6 del artículo 2º, Reglamento, 2013). Dentro de esta categoría se encuentran los datos biométricos que por sí solos permiten identificar al titular, datos de origen racial y étnico, ingresos económicos, opiniones o convicciones políticas, religión, afiliación sindical y toda aquella información vinculada a la salud o la vida sexual. De forma complementaria, el Reglamento de la Ley No. 29733, Ley de Protección de Datos Personales, (en adelante, "LPDP") (2011), (en adelante, "Reglamento", 2013, artículo 2.5) desarrolla el concepto de datos personales relacionados con la salud, referido a la información de la salud pasada (física y mental) incluyendo el estado pasado, presente y futuro de la persona, además del grado de discapacidad y su información genética. Respecto a la clasificación de los

> datos personales sensibles, se tiene una subclasificación: (i) aquellas reconocidas expresamente en la LPDP (2011) y su Reglamento (2013) y (ii) otras análogas que afecten su intimidad. Esta última constituye una especie de "caja de pandora" en la que consideramos que debe analizarse por el caso en específico. Por ello, un dato personal será sensible acorde a las circunstancias en las que se encuentre, pues, por ejemplo, no será lo mismo un correo electrónico de un usuario para afiliarse a un periódico digital (considerado como un dato personal), a que el mismo correo sea utilizado para acceder a contenido pornográfico, siendo este un dato personal sensible.

Asimismo, existe una parte de la doctrina que indica que la clasificación de la categoría de datos sensibles, en función a su naturaleza, constituye un error contraproducente, arbitrario e ineficiente, en tanto la clasificación (y protección) debe ir enfocada en la magnitud del daño o el riesgo de daño derivado de su recolección, uso o transferencia (Solove, 2024). Este problema se mitifica aún más en la era de la IA, donde algoritmos de aprendizaje automático facilitan las inferencias sobre datos sensibles a partir de datos no sensibles, llegando como resultado que – casi – todos los datos personales sean calificados como sensibles. Por ejemplo, la científica Sweeney (2000) identificó que, si se combinaban los datos de año de nacimiento, género y código postal, entonces había 87% de probabilidad de identificar a una persona.

De otro lado, Wacks (1993) elaboró una escala de sensibilidad, identificando tres niveles: alta (HS: *High sensitivity*), moderadeda (MS: *Moderate sensitivity*) y baja (LS: *Low sensitivity*). Según el autor, los datos altamente sensibles son aquellos íntimos, relacionados con información médica, sexual y otros aspectos de la vida del individuo. Los moderadamente sensibles son los que permiten juzgar a una persona y que si se utilizan inadecuadamente son potencialmente dañinos y los de baja sensibilidad son aquellos biográficos del individuo que facilitan la adquisición de información más sensible.

El autor expone que la creación de la tabla se originó considerando cómo la recopilación de dichos datos podría afectar al individuo. No obstante, se aclara que la tabla no es concluyente, ya que presenta una propuesta únicamente para garantizar protección y control legal, especialmente para los datos que lo necesitan con mayor énfasis.

Tabla 3.
Graduación de datos sensibles según *Wacks*

INFORMACIÓN	GRADO DE SENSIBILIDAD
Nombre	LS
Lugar de nacimiento	LS
Domicilio	LS
Nacionalidad, ciudadanía	LS
Estatus residencial	LS
Características físicas (altura, pelo, ojos)	LS
Sexo	LS
Número telefónico	LS
Ocupación	LS
Membresía a partido política (actual o previa)	MS
Membresía a organización política (actual o previa)	MS
Religión	MS
Membresía a una organización religiosa	MS
Investigaciones concernientes al individuo	MS
Antecedentes penales	MS
En juicio	MS
Buscado	MS
Sospechoso	MS
Estatus de inmigrante	MS
Raza	MS
Enfermedad venérea	HS

Fuente: Wacks (1993)

Según Wacks (1993), la presente tabla[11] fue diseñada a partir del alcance que la recolección de tales datos pudiese perjudicar al individuo. A nuestro juicio, la clasificación realizada por el autor no está conforme con la LPDP (2011) ni su Reglamento (2013), pues, por ejemplo, clasifica a los datos de membresía política o religión como moderadamente sensible cuando es, en realidad, un dato sensible relativamente alto, en tanto su aplicación negligente podría perjudicar al individuo por causa de discriminación, por ejemplo.

Sin perjuicio de la clasificación de Wacks (1993), el intento de la clasificación de datos sensibles viene desde tiempo atrás a través de los siguientes documentos internacionales:

Tabla 4.
Clasificación de datos sensibles en documentos internacionales

Cuerpos normativos	Definición de dato sensible
Directiva 95/46/CE del Consejo, 1995	Clasificaban a los datos sensibles como "*el tratamiento de datos personales que revelen el origen racial o étnico, las opiniones políticas, las convicciones religiosas o filosóficas, la afiliación sindical, así como el tratamiento de datos relativos a la salud o a la vida sexual*."

[11] Al final del cuadro se incluye un dato referente a la información médica de un paciente que, si bien no es relevante para la presente investigación, se cataloga como de alta sensibilidad por el autor permitiéndonos hacer una comparación entre los niveles de sensibilidad.

Anexo de las Recomendaciones del Consejo (23 de septiembre de 1980) de las Directrices que regulan la protección de la intimidad y los flujos transfronterizos de datos personales *(Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*)	Los Expertos debatieron los criterios de datos sensibles, en donde se no logró definir ningún conjunto de datos que se consideren universalmente considerados sensibles.

Siguiendo con las clasificaciones, otra importante a considerar es la establecida por la Dirección de Protección de Datos Personales (también denominada, "DPDP") dentro de la "sección 3: Formularios del TUPA" del Ministerio de Justicia y Derechos Humanos, que incluye el formulario de inscripción para bancos de datos personales de personas naturales y jurídicas que correspondan a la administración pública o privada[12]. En base a ello, se diferencian los siguientes tipos de datos personales:

- Datos de carácter identificativo: nombres y apellidos, No. DNI, No. RUC, No. de pasaporte, dirección del domicilio, teléfono, dirección de correo electrónico, imagen, voz, firma, firma electrónica, otros (detallar). Un ejemplo podría ser el carné de extranjería.
- Datos de características personales: estado civil, fecha de nacimiento, nacionalidad, sexo, profesión, edad, datos académicos, datos de derechohabientes, datos de persona de contacto, otros (detallar).
- Datos económicos-financieros y de seguros: créditos, préstamos, avales, datos bancarios, historial de créditos, información tributaria, seguros, tarjetas de crédito, bienes patrimoniales, planes de pensiones/jubilaciones, beneficios recibidos de programas sociales, hipotecas, deudas, otros (detallar).
- Datos de carácter social: pertenencia a clubes o asociaciones, aficiones y hábitos personales, características de vivienda, otros (detallar).

[12] Las denominaciones según la Dirección de Protección de Datos Personales son:
Formulario PDF: 012-A INSCRIPCIÓN DE BANCO DE DATOS PERSONALES DE ADMINISTRACIÓN PÚBLICA https://www.minjus.gob.pe/wp-content/uploads/2017/03/FORM-012-A.pdf
Formulario PDF: 012-B INSCRIPCIÓN DE BANCO DE DATOS PERSONALES DE ADMINISTRACIÓN PRIVADA-PERSONA JURÍDICA https://www.minjus.gob.pe/wp-content/uploads/2017/03/FORM-012-B.pdf
Formulario PDF: 012-C INSCRIPCIÓN DE BANCO DE DATOS PERSONALES DE ADMINISTRACIÓN PRIVADA-PERSONA NATURAL https://www.minjus.gob.pe/wp-content/uploads/2017/03/FORM-012-C.pdf

- Datos sensibles: origen étnico, convicciones filosóficas o morales, información relativa a la salud física o mental, afiliación sindical, vida sexual, convicciones religiosas, convicciones políticas, huella dactilar, características físicas, ingresos económicos, vida afectiva o familiar, otros datos de carácter biométrico (detallar).

Posteriormente, en el 2017 se reconoció la siguiente categoría de datos personales como un estándar internacional denominada ISO/IEC 19941 (2017) que se desarrolló para lograr la interoperabilidad y portabilidad en servicios de computación en la nube: (i) datos del cliente, es decir, datos aportados principalmente por un usuario de un proveedor de servicios en la nube (por ejemplo, credenciales, datos personales de salud y registros médicos, y financieros); ii) datos derivados, es decir, datos observados y/o inferidos sobre el usuario; iii) datos del proveedor de servicios en la nube, incluidos principalmente datos de operaciones y datos de acceso y autenticación; y iv) datos de la cuenta, incluyendo principalmente información de contacto de la cuenta o administración y datos del instrumento de pago. Dentro de esta clasificación, se distingue entre cinco categorías o estados de identificabilidad de los datos[13] (2018) y son los siguientes:

- Datos identificados: Datos que pueden asociarse inequívocamente con una persona específica porque en la información se puede observar información de identificación personal.
- Datos seudonimizados: Datos en los que todos los identificadores se sustituyen por alias cuya asignación de alias es tal que no puede revertirse con base a esfuerzos razonables por quien no los realizó.
- Datos seudonimizados desvinculados: Datos cuyos identificadores se borran o se sustituyen por alias cuya función de asignación es borrada o irreversible, de manera que el vínculo no puede restablecerse por ningún esfuerzo razonable de nadie incluida la parte que los realizó.
- Datos agregados: Datos estadísticos que no contienen entradas a nivel individual y se combinan a partir de información sobre suficientes personas diferentes para que los atributos a nivel individual no sean identificables.

[13] Se encuentran alineados con las técnicas de desidentificación que mejoran la privacidad incluidas en el estándar ISO/IEC 20889 (2018).

Una clasificación distinta es la ampliamente citada en el medio que fue elaborada por la OCDE (2019), la cual identifica ciertos tipos de datos personales según su forma de obtención:

- Datos ofrecidos voluntariamente (o entregados, aportados o proporcionados): por individuos cuando comparten explícitamente información sobre ellos mismos o sobre otros como, por ejemplo, al crear un perfil de red social y al ingresar información de tarjeta de crédito para comprar en línea.
- Datos observados: se crean donde se capturan y registran las actividades y el papel del interesado es pasivo, mientras que el responsable del tratamiento juega el papel activo. Por ejemplo, los datos de ubicación de teléfonos móviles o lo relacionado el comportamiento del uso de la web.
- Datos derivados (o inferidos o imputados): Según la OCDE (2014), se crean basándose en el análisis de datos e incluye los datos creados mecánicamente utilizando un razonamiento simple y matemáticas básica para detectar patrones. El procesador o encargado del tratamiento de datos tiene un rol activo. El sujeto de datos, normalmente, tiene poca conciencia sobre lo que se infiere sobre él o ella como, por ejemplo, los puntajes crediticios calculados en función del historial financiero de una persona. Asimismo, de acuerdo con Narayanan y Shmatikov (2006), en estos casos, la información personal puede derivarse de datos que, aparentemente, son anónimos o no personales.
- Datos adquiridos (comprados o con licencia): Datos obtenidos de terceros mediante contratos de licencia comercial. Por ejemplo, cuando los datos se adquieren de intermediarios de datos u otros medios no comerciales (por ejemplo, datos que tienen como fuente de obtención a las iniciativas de gobierno abierto). Así, tanto las obligaciones legales como contractuales pueden afectar la reutilización y el intercambio de dichos datos.

Adicionalmente, García (2019) refiere al almacenamiento de los datos personales y coincide con la legislación colombiana, tal como se señala a continuación:

- Datos personales públicos: Datos personales que son de conocimiento por un número cuantioso de personas, sin que el titular pueda saber, en todos los casos, la fuente o la forma de difusión de los mismos, ni que por la calidad de datos pueda impedir que, una

vez conocido, sea generalmente difundido en el marco de los límites de respeto y convivencia cívicos (Davara, 1993). Asimismo, es toda la información o datos dentro de documentos públicos, actos judiciales o administrativos que no se encuentren sometidos a reserva y datos del estado civil (Ley 1266, 2018).

- Datos personales privados: Es el dato que solo es relevante para su titular (Ley 1266, 2018) (Ejm. fotografías, videos, datos relacionados con su estilo de vida).
- Datos semiprivados: El dato semiprivado no cuenta con naturaleza pública, íntima o reservada; sin embargo, su divulgación o conocimiento es importante para un sector, grupo o la sociedad; por ejemplo, los datos crediticios o comerciales (Ley 1266, 2018). Son aquellos que constan en registros privados, por lo que puede haber tantos como cada titular desee realizar. Además de ser de interés para el titular, puede ser de interés para cierto sector o grupo de personas (Ejm. fecha y lugar de nacimiento).

Finalmente, los datos personales se pueden clasificar según su fuente[14] (Jervis, 2006):

- Datos directos: Aquellos que han sido entregados por el titular de forma directa y voluntaria, y cuya fuente u origen es la voluntad de su titular
- Datos indirectos: Datos que han sido recolectados o extraídos de o desde otros bancos de datos, a partir de su comunicación mediante acceso, cesión, transmisión, transferencia, interconexión u otro mecanismo análogo.

De lo revisado, concluimos que cada una de las instituciones sostiene una forma distinta de entender y clasificar los datos personales como tal. Sin perjuicio de ello, se desarrollará con mayor profundidad la normativa de protección de datos personales peruana en el acápite 2.4.

2.3 Desde la óptica constitucional peruana: la protección de datos personales

2.3.1 El derecho a la intimidad personal y familiar

Desde su derivación etimológica, el término íntimo procede del latino "*intimus*" que significa dentro. En palabras de Herrán (1998), la intimidad se refiere a todas las circunstancias internas del individuo que mantiene como núcleo de su personalidad. Por ello, lo íntimo constituye parte de su esencia individual que es propia y si ello fuera arrebatado, vería perjudicada la

[14] Esta clasificación cobra importancia en distintos aspectos, como, por ejemplo, en la necesidad y contenido del consentimiento del titular de los datos, en el requerimiento de notificación del tratamiento, entre otros supuestos equivalentes.

sustancialidad humana, su individualidad. Es por ello su importancia de reconocimiento como derecho fundamental.

Al respecto, sus antecedentes se remontan hacia finales del siglo XIX alrededor de la noción norteamericana del "*right of privacy*"[15] o derecho a la privacidad, que es el derecho de todo individuo de salvaguardar para sí aquellas facetas íntimas o familiares que así lo considere (Eguiguren, 2004a, p. 94). Asimismo, se define el derecho a la privacidad como "*the right to be let alone*", es decir, el derecho a ser dejado solo o sin ser molestado por intromisiones externas indeseables (como se citó en Morales, s.f.). Si bien es cierto que, a simple vista, se puede concluir que el derecho a la intimidad se asemeja al derecho a la privacidad o el llamado *right of privacy*; sin embargo, ambos conceptos, al día de hoy, no son análogos o equivalentes.

Para entender la definición del derecho a la intimidad, este último ha ido mutando a través del tiempo. Rivera Llano (1984) define a la intimidad personal como uno de los derechos de la personalidad en que cada uno encuentra la posibilidad de desarrollo y fomento de la personalidad en el exterior. Es decir, se trata de aquel territorio personal reservado a la curiosidad pública, absolutamente necesario para el desarrollo humano y donde enraíza la personalidad.

Aterrizando en el campo de los tratados sobre derechos humanos, el artículo 55 de la Constitución Política del Perú (1993), sostiene que los tratados forman parte del derecho interno, siendo que estos son expresiones de voluntad que adopta el Estado con organismos extranacionales y que se rigen por las normas, costumbres y fundamentos doctrinarios del derecho internacional (Tribunal Constitucional, 2006). Así, los tratados sobre derechos humanos tienen rango constitucional (Tribunal Constitucional, 2021).

Al respecto, existe una serie de tratados internacionales referentes al derecho a la intimidad:

- Declaración Universal de los Derechos Humanos (Asamblea General de las Naciones Unidas, 1948): el artículo 12 establece que nadie será objeto de injerencias arbitrarias en su vida privada, su familia, su domicilio o su correspondencia, ni de ataques a su honra o a su reputación.

[15] El término right to privacy fue formulado en 1890 por Warren, D. y Brandeis, L.D. en su célebre artículo The right to privacy, publicado en "Harvard Law Review".

- Declaración Americana de los Derechos y Deberes del Hombre (Organización de los Estados Americanos, 1948): el artículo V indica que constituye un derecho de la persona la protección contra los ataques desmesurados en perjuicio de su honra, reputación, vida íntima y familiar.
- Pacto Internacional de Derechos Civiles y Políticos (Naciones Unidas, 1966): el artículo 17 dispone que es un derecho de la persona la protección contra las injerencias o ataques sobre su esfera íntima, familia, domicilio o su correspondencia y su honra y reputación.
- Convención Americana sobre Derechos Humanos - Pacto de San José (Organización de los Estados Americanos, 1969): al igual que los demás tratados, se encuentra redactado en la misma línea dentro del artículo 11[16].

El derecho a la intimidad se encuentra inmediatamente enlazado con la dignidad de la persona, tanto a nivel constitucional, como legislativo y jurisprudencial (como se citó en Lucas, 1995). Distintos juristas sostienen que, además del Derecho, se encuentra regulado dentro del ámbito de la ética (González, 1990). Así pues, siendo que el derecho a la intimidad se encuentra inseparablemente unido a la dignidad, entonces se encuentra relacionado con los derechos de la personalidad, siendo que en este último se distinguen tres esferas: la intimidad, vida privada y pública (Desantes, 1990). Así, el derecho a la intimidad y vida privada involucra al conjunto de actos, situaciones o circunstancias que no se encuentran normalmente expuestos ante el dominio público, por lo que se incluye a la libertad del sujeto en desarrollarse sin ningún tipo de particularidades ocasionadas por terceros o por la extensión de hechos privados (Eguiguren, 2002).

En esa línea, el derecho a la intimidad cuenta con dos dimensiones: (i) como secreto de la vida privada y (ii) como libertad individual. Por un lado (i) atentan contra la intimidad todas aquellas divulgaciones ilegítimas de hechos que se relacionan tanto con la vida privada, familiar o las investigaciones ilegítimas de acontecimientos propios de dicha vida. Por otro lado (ii), la intimidad trasciende y se realiza en el derecho de toda persona a tomar por sí misma decisiones

[16] Convención Americana sobre Derechos Humanos,
Artículo 11. Protección de la Honra y de la Dignidad
1. Toda persona tiene derecho al respeto de su honra y al reconocimiento de su dignidad.
2. Nadie puede ser objeto de injerencias arbitrarias o abusivas en su vida privada, en la de su familia, en su domicilio o en su correspondencia, ni de ataques ilegales a su honra o reputación.
3. Toda persona tiene derecho a la protección de la ley contra esas injerencias o esos ataques.

que impactan y conciernen a la esfera de su vida privada (Comisión Andina de Juristas, 1997). Podemos apreciar entonces que este derecho aparece configurado como un derecho de doble perspectiva: (a) una negativa, que se traduce en el poder de exclusión del conocimiento ajeno de aquello que se refiere a la propia persona, y (b) una perspectiva positiva, de vigilancia y control del individuo sobre la información que le concierne (Herrán, 2002).

Remitiéndonos a nuestra jurisdicción local, la intimidad es un derecho fundamental reconocido en el artículo 2.7 de la Constitución (1993), indicando que toda persona tiene derecho a la intimidad personal y familiar[17]. Además, existen otros dispositivos normativos alojados en la Constitución (1993) que se encuentran alineados al citado artículo 2.7, como es la no remisión de información, por parte de servicios informáticos (artículo 2.6).

Resulta importante resaltar el artículo 12 de nuestra Carta Magna, en donde se establece que ninguna persona estará sujeta a interferencias arbitrarias en su vida privada, familia, domicilio o correspondencia. En virtud de ello, se reconoce el derecho a la protección legal frente a tales intervenciones o amenazas. De este modo, en el marco constitucional, en caso existiera ciertas divergencias de conceptos entre la legislación nacional e internacional, ello se unifica a través de la Cuarta Disposición Final y Transitoria de la Constitución señalado anteriormente y del artículo V del Código Procesal Constitucional.

El Tribunal Constitucional ha advertido que la intimidad personal implica el aislamiento de la intromisión de terceros de todos aquellos aspectos de la persona que forman parte de su desarrollo interno, entendido como el desarrollo de su personalidad física y espiritual que se encuentra reservada para sí misma, entre los que hallamos el desarrollo de los procesos de pensamiento y opinión, de la salud física y emocional, de la sexualidad humana (en todas sus expresiones), entre otros aspectos que únicamente son de interés de la persona (Tribunal Constitucional, 2014, fundamento 12).

Respecto a la vida privada, esta se define como un conjunto de acciones y relaciones interpersonales que se distingue por ser de carácter restringido y personalísimo, en donde a su vez involucran afectivamente a las personas y las proyectan conjuntamente por medio de una

[17] Constitución Política del Perú
Artículo 2.- Toda persona tiene derecho:
7. Al honor y a la buena reputación, a la intimidad personal y familiar, así como a la voz y a la imagen propias.

red de compromisos, disponibilidad, responsabilidades, cuidados y expectativas (Álvarez, 2021). De acuerdo a Eguiguren (2002), ello puede considerarse la facultad de todo ciudadano a reservarse una esfera para evitar su manipulación o instrumentalización, excepto los personajes públicos que, por interés general, deciden reducir los márgenes de su vida privada o su intimidad, sin perjuicio de que se respete el derecho al honor y a la vida privada.

En cuanto a nuestro Tribunal Constitucional (2007), consideramos oportuna la definición propuesta al señalar que el derecho a la vida privada constituye aquellos hechos, datos o situaciones no conocidas por el exterior y que, siendo verídicos, se encuentran en reserva por parte del sujeto o de un grupo de terceros reducido, considerando que su divulgación puede traer consigo algún daño. Asimismo, se considera que la vida privada implica necesariamente la posibilidad de excluir a los demás, y que, resulta indispensable para la realización del ser humano, a través del libre desarrollo de su personalidad, de conformidad con el artículo 2° inciso 1 de la Constitución (Tribunal Constitucional, 2005).

Por otro lado, nuestro Tribunal Constitucional ha determinado que el secreto bancario y reserva tributaria no se encuentran inmersos dentro de la esfera de la intimidad y vida privada (Tribunal Constitucional, 2004a). Esto porque solo están constitucionalmente prohibidas aquellas restricciones cuya intención sea vulnerar la esfera privada del individuo y causarle perjuicios reales y/o potenciales de diversas índoles, por lo que no se incluyen aquellas limitaciones que, manteniendo el carácter reservado inherente, permitan alcanzar objetivos constitucionalmente legítimos y se ubiquen dentro de los límites establecidos por la razonabilidad y la proporcionalidad (Tribunal Constitucional, 2011, f. 11-14).

De todo lo señalado previamente, podemos considerar al derecho a la intimidad personal y vida privada como positiva y amplia, en tanto nuestro máximo intérprete constitucional considera que es una plataforma para el desarrollo e integración del ser humano, en cuyo caso abarca esferas físicas, familiares y espirituales, sea en cualquier tipo de comunicación (Tribunal Constitucional, 2005, f. 30). Además, el Tribunal Constitucional delimita que la intimidad, a pesar de su carácter personalísimo (aplicable a personas físicas), se le reconoce también (en algunos casos) a las personas jurídicas, lo que se condice con la doctrina del Derecho comparado[18].

[18] El Tribunal Constitucional, a través de la Sentencia recaída en el Expediente No. 0004- 2004-AI (f.j. 34 y ss.). ha confirmado la doctrina propuesta en la Sentencia recaída en el Expediente No. 1219- 2003-HD (f.j. 9) según la cual, "el secreto bancario

Respecto al artículo 14 del Código Civil (1984), si bien dicho cuerpo normativo dedica varios artículos en defensa de los derechos de intimidad cuando se producen injerencias ilegítimas en tales ámbitos protegidos por la ley (Rodríguez, 2011), no se adiciona algún contenido complementario que nos permita determinar su contenido y alcances. A pesar de este vacío normativo, consideramos que tal derecho no debe ser interpretado de manera limitativa, sino más bien aperturada incluyendo a todo acto de vigilancia o captación por parte del Estado o un particular que vulnere el derecho a la intimidad de una persona.

2.3.2 El derecho a la autodeterminación informativa

Si bien desde una primera (y ligera) interpretación el derecho a la intimidad personal y a la vida privada puede ser confundido con el derecho a la autodeterminación informativa, en tanto se ha señalado erróneamente que este último constituye una ampliación del derecho fundamental a la intimidad (Rebollo, 2003); ello resulta equívoco una vez analizado el contenido intrínseco y radiográfico del derecho a la autodeterminación informativa.

En principio, los derechos fundamentales se basan en la dignidad humana, lo cual está plenamente reconocido en nuestra Carta Magna como pilar fundamental[19]. Siendo ello así, la autodeterminación informativa es aquel derecho que goza una persona de poder controlar sus datos personales, información de índole personal y familiar, así como toda aquella que no cuente con dichas características, pero que de igual forma sea objeto de control y reserva (Eguiguren, 2004b). No se trata de la expresión del contenido de otro derecho ni el resultado de la combinación o más ya declarados en la Constitución, sino de un pleno reconocimiento expreso constitucional autónomo (Murillo y Piñar, 1990, p. 71).

Este derecho cuenta con dos dimensiones: (i) libertad negativa al prohibir la injerencia o intrusión de terceros y una (ii) libertad positiva, en tanto el sujeto de derecho tendría el pleno control de todos los aspectos considerados íntimos (Murillo y Piñar, 1990, p. 152). En esa línea, la autodeterminación informativa no significa crear algún tipo de "propiedad" respecto de la

forma parte del contenido constitucionalmente protegido del derecho a la intimidad, y su titular es siempre el individuo o la persona jurídica de derecho privado que realiza tales operaciones bancarias o financieras".

[19] Constitución Política del Perú
Artículo 1.- La defensa de la persona humana y el respeto de su dignidad son el fin supremo de la sociedad y del Estado.

información, sino es imponer al Estado la tarea de organizar el tratamiento de datos de tal manera que se asegure el respeto de la autonomía de los interesados (León, 2006).

La autodeterminación informativa, en palabras de Adinolfi (2006, p. 132), es la posibilidad de autorizar, de bloquear, de oponerse, de ratificar, de quedarse indiferente respecto a las circulaciones de voces, *rectius* informaciones, acerca de la persona misma. En base a ello, este derecho permite (i) la reducción del volumen de información manejada; (ii) evitar la generación de nuevos elementos informativos a raíz del tratamiento de datos; (iii) evitar que se difunda información íntima, (iv) evitar la elaboración de perfiles de la personalidad para la toma de decisiones y (v) evitar la comercialización de dichos resultados.

Asimismo, es importante señalar que el derecho a la autodeterminación informativa tiene un enlace pleno y sólido con la privacidad; sin embargo, no constituyen lo mismo. El término privacidad se deriva de lo privado (Flores, 1987, p. 435). Lo conforman aquellas acciones propias, particulares y personales de los individuos, correspondiéndole solo al titular decidir sobre ellos y constituye uno de los valores más importantes de respeto al ser humano (Quiroz Papa de García, 2016). De ese modo, el derecho a la autodeterminación informativa o derecho a la protección de datos personales tiene un carácter autónomo y subjetivo, que tiene sus raíces en la dignidad humana, abarcando el derecho a la intimidad, mas no siendo el mismo derecho.

Con el fin de entender el derecho de la autodeterminación informativa en el plano nacional, resulta imprescindible remitirnos a la legislación española para entender el nacimiento de este derecho en nuestra legislación. España reconoció, a través de su Constitución (1978), el límite en el uso de herramientas informáticas que protejan la intimidad familiar y personal de la persona (Prego, 2017). Distintos constitucionalistas españoles consideran que esta disposición fue el punto de partida para reconocer el derecho fundamental a la autodeterminación informativa que permite preservar la identidad al dominar la revelación y la usabilidad de los datos, protegiendo además de la ilimitada capacidad de archivar, relacionar y transmitir tales datos de los peligros existentes (León, 2006). Sin embargo, se puede identificar que el bien jurídico protegido desborda al derecho a la intimidad, generando la configuración de un derecho fundamental *siu generis*, es decir, el derecho a la autodeterminación informativa o derecho a la protección de datos personales.

Esta disposición fue importada de la jurisprudencia alemana en 1982 por un caso de una ley de censo popular (*Recht auf informationelle Selbstsbestimmung*), por el cual las autoridades alemanas resolvieron que la libre personalidad presupone que la defensa del individuo sea contra el acopio, almacenaje, empleo y transmisión ilimitada de los datos de la persona[20]. De ese modo, se ratificó que lo protegido no es la intimidad en sentido pleno, sino la libertad del individuo de decidir los límites sobre qué (y no) puede revelar, es decir, tener control sobre los mismos.

Años más tarde, se publicó la Ley Orgánica del Tratamiento Automatizado de Datos de Carácter Personal (Ley Orgánica 5/1992), cuya exposición de motivos sostiene expresamente que el artículo 18.4 de la Constitución Española refleja una protección a la privacidad, mas no la intimidad, en tanto la privacidad es más amplia al proteger el ámbito del desarrollo de la persona (por ejemplo, su domicilio, infancia, vida académica, hábitos de vida y consumo, vida profesional o laboral, creencias religiosas e ideologías[21]) que merece reserva plena (Ley Orgánica 5/1992,1992).

Asimismo, esta protección se basa en que, anteriormente, la privacidad estaba limitada por el (i) tiempo, pues en su transcurso tales recuerdos o hechos desaparecían, y (ii) por la distancia, en tanto no se podía conocer más allá de lo que ocurría alrededor. Considerando el avance de las tecnologías emergentes que permiten el acceso, recolección y almacenaje de tales datos es que tales limitantes prácticamente han desaparecido y que, por tanto, es el objeto de protección del artículo 18.4 de la Constitución Española (1978).

Si bien el citado artículo 18.4 se encuentra limitado al únicamente inferirse a la intimidad personal y familiar, el Tribunal Constitucional Español (2000) ha emitido distintas sentencias reconociendo la autonomía e independencia de la autodeterminación informativa frente al citado derecho. Asimismo, España ha firmado la Carta de los Derechos Fundamentales de la Unión Europea (2009), cuyo contenido protege el derecho a la protección de datos personales[22].

[20] Actualmente, la *Bundesdatenschutzgesetz* (Ley Federal de Protección de Datos) sostiene que la finalidad de la norma es proteger al individuo para que su derecho de la personalidad no resulte perjudicado por el negligente tratamiento de datos.
[21] Es importante señalar que el artículo 16.2 de la Constitución Española (1978) señala que "*nadie podrá ser obligado a declarar sobre su ideología, religión o creencias*"; es decir, se reconoce un tipo de dato sensible en vía constitucional.
[22] Carta de los Derechos Fundamentales de la Unión Europea
Artículo 8 Protección de datos de carácter personal
1. Toda persona tiene derecho a la protección de los datos de carÆcter personal que la conciernan.
2. Estos datos se tratarÆn de modo leal, para fines concretos y sobre la base del consentimiento de la persona afectada o en virtud de otro fundamento legítimo previsto por la ley. Toda persona tiene derecho a acceder a los datos recogidos que la conciernan y a su rectificación.

Es así que dicha Carta constituye un elemento ejemplar para el Tribunal Español, en tanto cada vez más se encuentra el reconocimiento fundamental de la protección de datos personales como un derecho autónomo.

Del mismo modo, el Tribunal Constitucional Español (1993) reconoce la autodeterminación informativa como "libertad informática", entendido como aquel derecho a controlar el uso de los datos insertos en un programa informático (habeas data) e incluye la oposición del individuo frente a terceros que utilizan sus datos personales incluso para fines que no otorgó autorización (Tribunal Constitucional Español, 2000). Como sostiene Fernández (1997), es el derecho de disponer, preservar, consentir, controlar y, en su caso, rectificar los datos informativos concernientes a la propia personalidad.

La Declaración Universal de los Derechos Humanos, ratificado por el Perú en 1959, sostiene que toda persona tiene derecho a la protección sobre injerencias arbitrarias de su vida privada[23]. Adicionalmente, la Declaración Americana de los Derechos y Deberes del Hombre (1948) sostiene que toda persona tiene derecho a la protección de la ley contra su vida privada[24].

Sobre el particular, no existe un instrumento en el derecho internacional sobre derechos humanos que reconozca expresamente el derecho a la protección de los datos personales en sistemas computarizados o informáticos, pues únicamente se circunscriben a la protección de la intimidad personal del individuo[25]. Sin perjuicio de ello, no significa que dicho derecho se encuentre desprotegido de por sí, pues la legislación local o interna regula cada uno de los parámetros aplicables.

La Constitución Política del Perú de 1979 fue la primera en reconocer el derecho a la intimidad personal y familiar. Sin embargo, únicamente se limitaba a ello, pues no señalaba un inciso específico referido a velar por la intimidad de la persona en los servicios informáticos o

3. El respeto de estas normas quedará sujeto al control de una autoridad independiente.

[23] Declaración Universal de los Derechos Humanos
Artículo 12.-
Nadie será objeto de injerencias arbitrarias en su vida privada, su familia, su domicilio o su correspondencia, ni de ataques a su honra o a su reputación. Toda persona tiene derecho a la protección de la ley contra tales injerencias o ataques

[24] Declaración Americana de los Derechos y Deberes del Hombre
Artículo V. Toda persona tiene derecho a la protección de la Ley contra los ataques abusivos a su honra, a su reputación y a su vida privada y familiar.

[25] Resaltamos el Convenio 108 del Consejo de Europa "Para la protección de las personas con respecto al tratamiento automatizado de datos de carácter personal", aprobado el 28 de enero de 1981; y el artículo 45 de las "Directrices para la regulación de ficheros automáticos de datos personales" de las Naciones Unidas. Sin embargo, ambos cuerpos jurídicos no constituyen tratados sobre derechos humanos.

máquinas computarizadas. Fue a partir de la Constitución (1993) que reconoce, dentro del artículo 2.6, la protección de la intimidad dentro de sistemas informáticos y aquellos que no[26].

Durante el debate de la Constitucional Pleno de 1993, en el que distintos juristas se reunieron para dialogar, analizar y debatir la elaboración de la Constitución vigente, es que se presenta la participación del Dr. Torres y Torres Lara, quien señala expresamente que la protección que debe dar la Constitución es que no se pueda transferir informaciones sobre la intimidad personal por medios comunes, siendo que es mucho más dañino comunicando una información negativa sobre una persona que acumulándola (Congreso Constituyente Democrático, 1998).

Sin perjuicio del análisis normativo que tuvieron los legisladores para la redacción de nuestra Carta Magna, el artículo 2.6 de la Constitución (1993) resulta deficiente y hasta cierto punto limitante. En primer lugar, el citado artículo 2 inciso 6 no reconoce la protección expresa del derecho a la autodeterminación informativa, ya que únicamente se limita a responder por la intimidad personal y familiar. Esto implica que no se pueda dilucidar como tal el carácter autónomo del derecho a la autodeterminación informativa.

No obstante, el derecho a la autodeterminación informativa no puede identificarse con el derecho a la intimidad, personal o familiar, reconocido, a su vez, por el inciso 7) del mismo artículo 2° de la Constitución. Es importante recordar que, mientras que el derecho a la intimidad protege el derecho a la vida privada, es decir, el poder jurídico de rechazar intromisiones ilegítimas en la vida íntima o familiar de las personas; el derecho a la autodeterminación informativa garantiza la facultad de todo individuo de poder preservarla controlando el registro, uso y revelación de los datos que les conciernen (Tribunal Constitucional, 2003, f.3). De hecho, el propio Tribunal Constitucional ha señalado expresamente que el derecho a la autodeterminación no puede identificarse con el derecho a la intimidad personal o familiar (Tribunal Constitucional, 2007a), ni con el derecho a la imagen (Tribunal Constitucional, 2003), ni con el derecho al acceso a la información pública (Tribunal Constitucional, 2018).

[26] Constitución Política del Perú
Artículo 2.- Toda persona tiene derecho:
6. A que los servicios informáticos, computarizados o no, públicos o privados, no suministren informaciones que afecten la intimidad personal y familiar.

En segundo lugar, considerando que el razonamiento del legislador fue reconocer el derecho de la autodeterminación informativa, este igualmente fue limitado al únicamente regular una parte del mismo. Ello en tanto no solo involucra la mera comunicación de información, sino también el acceso, rectificación (o aclaración), actualización y eliminación de dichos datos. De acuerdo a Luna (2021, p. 256), el derecho a la autodeterminación informativa incluye (i) si se van a exponer los datos personales, se realice con fidelidad, sin distorsionar las situaciones que los vinculan a ciertos hechos, y hasta donde sea lícito y razonable hacerlo; (ii) si va a usar los datos personales, que se realice bajo consentimiento, de manera proporcional y para la finalidad declarada y conocida por la persona; (iii) si se van a transferir los datos personales, que se conozca a dónde y a quiénes; (iv) si va a almacenar y procesar los datos personales, que se haga con seguridad e informando adecuadamente junto a la autoridad competente y (v) se realice el tratamiento observando las prescripciones legales previstas. Bajo la lógica del legislador, en caso quisiéramos realizar una interpretación estricta y literal del citado artículo, una persona no podría acceder a su información, así como tampoco suprimir, actualizar o solicitar su eliminación.

En tercer lugar, no existe una referencia expresa sobre qué tipo de datos se encuentran protegidos por medio del citado artículo de la Constitución (1993). La redacción del artículo sostiene que se trata de datos relacionados con la intimidad personal y familiar; sin embargo, ¿dónde quedan los demás tipos de datos que también merecen protección? Por ejemplo, aquellos vinculados con la vida privada que merecen una naturaleza de reserva, pero que no son estrictamente "íntimos", como, por ejemplo, información relacionada al desempeño profesional, entre otros.

En cuarto lugar, el artículo únicamente se limita a indicar sobre un "servicio informático", lo cual se puede entender como aquellas entidades públicas y/o privadas responsables de la recopilación, organización, almacenamiento, comunicación por transferencia de los datos personales; con fines de brindar un servicio general al público. Sin embargo, tal redacción queda nuevamente limitante y restrictiva, pues el tratamiento de los datos personales que puedan realizar las entidades públicas quedaría exceptuado del ámbito de aplicación del referido artículo, así como todas aquellas entidades privadas que no realizan servicios informáticos, pero sí realizan tratamiento de datos personales.

En vista del análisis exhaustivo del artículo 2.6 de la Constitución (1993), a nuestro juicio, tal redacción que, si bien tiene cierto grado de inspiración e influencia de la legislación española, queda limitada e incompleta, trayendo como consecuencia una afectación en la aplicación e interpretación en la práctica. De acuerdo con León (2011), la vía correcta de reconocer este derecho sería con el artículo 3 de la Constitución (1993)[27], dado el carácter restrictivo y altamente limitado del artículo 2 inciso 6.

Sin perjuicio de ello, nuestro Tribunal Constitucional, considerado el órgano supremo intérprete de la Constitución y que goza del control constitucional, ha establecido, en distintos pronunciamientos, que el derecho a la autodeterminación informativa se encuentra resguardado en el artículo 2.6 de la Constitución. Ello se puede ratificar en las siguientes sentencias:

Tabla 5.
Principales sentencias en materia de autodeterminación informativa

Sentencia del Exp. No. 1797 – 2002-HD/TC (2003)	Constituye la primera sentencia en reconocer expresamente que el derecho de la autodeterminación informativa se encuentra protegida conforme con el artículo 2.6 de la Constitución. Advierte que existe diferencia con el derecho a la identidad personal, en tanto este último corresponde al derecho a que la proyección social de la propia personalidad no sufra interferencias o distorsiones a causa de la atribución de ideas, opiniones, o comportamientos diferentes de aquellos que el individuo manifiesta en su vida en sociedad.
Sentencia del Exp. No. 4739-2007-HD (2007b)	Reconoce que el derecho a la autodeterminación informativa consiste en la serie de facultades que tiene toda persona para ejercer control sobre la información personal que le concierne, contenida en registros ya sean públicos, privados o informáticos. De ese modo, protege al titular de la información de posibles abusos o riesgos derivados de la utilización de datos personales contenidos en registros ya sean públicos o privados.
Sentencia del Exp. No. 00746-2021 O-PHD/TC (2022)	Indica que la autodeterminación informativa también garantiza que una persona pueda hacer uso de la información privada que existe sobre él, ya sea que ésta se encuentre almacenada o en disposición de entidades públicas o de carácter privado.
Sentencia del Exp. No. 03100-	Señala que la Constitución reconoce como derecho fundamental la autodeterminación informativa. Indica, además, que, en el marco del ejercicio del derecho a la autodeterminación informativa, una persona puede solicitar ante cualquier entidad, sea pública o privada,

[27] Constitución Política del Perú
Artículo 3.- La enumeración de los derechos establecidos en este capítulo no excluye los demás que la Constitución garantiza, ni otros de naturaleza análoga o que se fundan en la dignidad del hombre, o en los principios de soberanía del pueblo, del Estado democrático de derecho y de la forma republicana de gobierno.

2021-PHD/TC (2022)	información creada en torno a la actividad que realiza –o realizó en su respectivo momento.

Fuente: Elaboración propia

Sin lugar a dudas podemos señalar que, para nuestro legislador, el derecho a la autodeterminación informativa (o derecho a la protección a los datos personales) se encuentra plenamente reconocido en nuestra legislación nacional. Por ello, resaltamos que la finalidad del derecho de la autodeterminación informativa es asegurar protección e imponer limitaciones y control frente al recojo, almacenamiento, sistematización, elaboración, transmisión o difusión de datos personales que realizan instituciones públicas o entidades privadas sea (o no) un servicio computarizado (Eguiguren, 2004), sin perjuicio de que se traten de registros públicos o privados (Tribunal Constitucional, 2022). Así, se trata de un derecho subjetivo que es de naturaleza relacional al estar vinculado, muchas veces, con la protección de otros derechos constitucionales (Tribunal Constitucional, 2003) y que las condiciones de almacenamiento (o, correctamente, el tratamiento) de los datos personales deben cumplir con los criterios de veracidad, integridad, utilidad y caducidad (Tribunal Constitucional, 2022).

2.3.2.1 El proceso de *Habeas Data*

En vista de las posibles amenazas y peligros que existen y afectan al derecho a la autodeterminación informativa, además de optar por la vía regulatoria a través de un procedimiento trilateral – explicado en el acápite 2.4-, el legislador peruano constituyó un mecanismo o vía para resguardar este derecho dentro de la Constitución (1993) la que estableció, en su artículo 200°, inc. 3, dentro del Título que regula las "Garantías Constitucionales". Es así que el proceso constitucional de *Habeas Data* procede contra el hecho u omisión, por parte de cualquier autoridad, funcionario o persona, que vulnera o amenaza los siguientes derechos: a solicitar y obtener información de entidades públicas (Constitución, 1993, Art. 2°, inc. 5°); a que los servicios informáticos -públicos o privados- no suministren informaciones que afecten la intimidad personal y familiar (Constitución, 1993, Art 2.6)[28].

[28] Constitución Política del Perú
Artículo 200.- Son garantías constitucionales:
3. La Acción de Hábeas Data, que procede contra el hecho u omisión, por parte de cualquier autoridad, funcionario o persona, que vulnera o amenaza los derechos a que se refiere el artículo 2, incisos 5 y 6 de la Constitución.

Desde su reconocimiento constitucional, existieron opiniones divididas sobre el particular, en tanto consideraban que la inclusión de tal proceso constitucional especializado resultaría innecesaria, pues contendría igual efecto que el proceso de amparo (Abad, 1994, p. 268). Asimismo, la doctrina autorizada consideró una desnaturalización del *Habeas Data* que incluya el derecho de rectificación (García Belaunde, 1994), pues este se protege a través de la vía administrativa.

Sin contar ello, el proceso de *Habeas Data* es conocido por el juez constitucional del lugar donde se encuentre la información, el dato o donde tiene su domicilio principal el afectado, a elección del demandante (Nuevo Código Procesal Constitucional [NCPC], 2021)[29]. Además, solo podrá interponerlo el afectado, sus tutores o curadores o por sus herederos y, en caso se trate de una persona jurídica, por su representante legal (NCPC, 2021)[30].

Un requisito relevante e imprescindible a considerar dentro de la demanda es que el afectado deberá exponer en su demanda las razones por las cuales se entiende que en el archivo, registro o banco de datos individualizado obra información referida al agraviado; los motivos por los cuales considera que la información que le atañe resulta discriminatoria, falsa, inexacta o violatoria de la intimidad personal o familiar (NCPC, artículo 57)[31].

Por su parte, EL NCPC dispone expresamente que el *Habeas Data* procede en defensa del derecho a la autodeterminación informativa, enunciando dieciséis (16) modalidades sobre el referido derecho (NCPC, artículo 59)[32]. La finalidad del NCPC se enmarca en volver las cosas

[29] Ley No. 31307, Nuevo Código Procesal Constitucional
Artículo 54. Juez competente
Es competente para conocer los procesos de habeas data, el juez constitucional del lugar donde se encuentre la información, el dato o donde tiene su domicilio principal el afectado, a elección del demandante. En el proceso de habeas data, no se admitirá la prórroga de la competencia territorial, bajo sanción de nulidad de todo lo actuado.
[30] Ley No. 31307, Nuevo Código Procesal Constitucional
Artículo 55.- Legitimación activa
La demanda de habeas data solo puede ser ejercida por el afectado, sus tutores o curadores o por sus herederos. Cuando la demanda es interpuesta por persona jurídica de derecho privado, esta se interpone por su representante legal o por el apoderado que designe para tal efecto
[31] Ley No. 31307, Nuevo Código Procesal Constitucional
Artículo 57.- Requisitos especiales de la demanda de habeas data
Además de los requisitos establecidos en el artículo 2, la demanda de habeas data contiene:
1. El nombre y domicilio del archivo, registro o banco de datos y, en su caso, el nombre del responsable o usuario. En caso de los archivos, registros o bancos públicos, se procurará establecer el organismo estatal del cual dependen.
2. Las razones por las cuales se entiende que en el archivo, registro o banco de datos individualizado obra información referida al agraviado; los motivos por los cuales considera que la información que le atañe resulta discriminatoria, falsa, inexacta o violatoria de la intimidad personal o familiar.
[32] Ley No. 31307, Nuevo Código Procesal Constitucional
Artículo 59.- Derechos protegidos

al estado anterior a la violación de un derecho constitucional o disponiendo el cumplimiento de un mandato legal.

La doctrina (Chiriboga, 2001) sostiene que el *Habeas Data* es una garantía constitucional con objetivos precisos, pues busca que el accionante conozca:

- Los motivos legales por los cuales el poseedor de la información llegó a obtenerla.
- Desde cuándo se cuenta con la información.
- El uso y lo que se hará con la información.
- Qué personas (naturales o jurídicas) conocieron la información por el poseedor, incluyendo el motivo de su transmisión, propósito y fecha.
- Las tecnologías que usan para almacenar la información.
- Las medidas de seguridad que tiene el poseedor.
- Qué información tiene en su poder.
- Conocer si la información es correcta y actualizada, de lo contrario, solicitar la actualización o rectificación de dicha información.

De acuerdo con el Tribunal Constitucional (1998), el derecho a la autodeterminación informativa a través del *Hábeas Data* comprende, en primer lugar, la capacidad de exigir

El habeas data procede en defensa del derecho de acceso a la información pública reconocido en el inciso 5) del artículo 2 de la Constitución. También procede en defensa del derecho a la autodeterminación informativa, enunciativamente, bajo las siguientes modalidades:
1) Reparar agresiones contra la manipulación de datos personalísimos almacenados en bancos de información computarizados o no.
2) A conocer y supervisar la forma en que la información personal viene siendo utilizada.
3) A conocer el contenido de la información personal que se almacena en el banco de datos.
4) A conocer el nombre de la persona que proporcionó el dato.
5) A esclarecer los motivos que han llevado a la creación de la base de datos.
6) A conocer el lugar donde se almacena el dato, con la finalidad de que la persona pueda ejercer su derecho.
7) A modificar la información contenida en el banco de datos, si se trata de información falsa, desactualizada o imprecisa.
8) A incorporar en el banco de datos información que tengan como finalidad adicionar una información cierta pero que por el transcurso del tiempo ha sufrido modificaciones.
9) A incorporar información que tiene como objeto aclarar la certeza de un dato que ha sido mal interpretado.
10) A incorporar al banco de datos una información omitida que perjudica a la persona.
11) A eliminar de los bancos de datos información sensible que afectan la intimidad personal, familiar o cualquier otro derecho fundamental de la persona.
12) A impedir que las personas no autorizadas accedan a una información que ha sido calificada como reservada.
13) A que el dato se guarde bajo un código que solo pueda ser descifrado por quien está autorizado para hacerlo.
14) A impedir la manipulación o publicación del dato en el marco de un proceso, con la finalidad de asegurar la eficacia del derecho a protegerse.
15) A solicitar el control técnico con la finalidad de determinar si el sistema informativo, computarizado o no, garantiza la confidencialidad y las condiciones mínimas de seguridad de los datos y su utilización de acuerdo con la finalidad para la cual han sido almacenados.
16) A impugnar las valoraciones o conclusiones a las que llega el que analiza la información personal almacenada.

jurisdiccionalmente la posibilidad de acceder a los registros de información, computarizados o no, cualquiera que sea su naturaleza, en los que se encuentren almacenados los datos de una persona. Tal acceso puede tener por objeto que se permita conocer qué es lo que se encuentra registrado, para qué y para quién se realizó el registro de información y la (o las) persona(s) que recabaron dicha información. En segundo lugar, el *Hábeas Data* puede agregar datos al registro que se tenga, ya sea por la necesidad de que se actualicen los que se encuentran registrados, o que se incluyan aquellos no registrados, pero que son necesarios para tener una cabal referencia sobre la imagen e identidad de la persona afectada. Asimismo, mediante el *Hábeas Data*, un individuo puede rectificar la información, personal o familiar, que se haya registrado; impedir que esta se difunda para fines distintos de los que justificaron su registro o, incluso, tiene la potestad de cancelar aquellos que razonablemente no debieran encontrarse almacenados.

Antes de la entrada en vigencia del citado NCPC, el Alto Tribunal (2007b) realizó una interesante clasificación sobre la naturaleza del *Habeas Data*, siendo que los dividió en *Habeas Data* puro y *Habeas Data* impuro.

El *Habeas Data* puro corresponde a reparar agresiones contra la manipulación de datos personalísimos almacenados en bancos de información computarizados o no. Estos a su vez se dividen en los siguientes:

- *Habeas Data* de Cognición: se trata de efectuar una tarea de conocimiento y de supervisión sobre la forma en que la información personal almacenada está siendo utilizada. Estos pueden ser Informativos (qué dato se guarda), Inquisitivo (quién proporcionó el dato), Teleológico (para qué se creó el dato personal) y de Ubicación (dónde está ubicado el dato). Por ejemplo, este tipo de *Habeas Data* cumple con el artículo 18 de la LPDP (2011).
- *Habeas Data* Manipulador: tiene como propósito la modificación de la información almacenada. Puede ser Aditivo (implica la actualización, aclaración o adición), Correctivo (modificar), Supresorio (eliminar), Confidencial (acceso restringido), Desvinculador (impedir que terceros tomen conocimiento), Cifrador (guardado en código), Cautelar (impedir la manipulación o publicación del dato), Garantista (control

técnico de los datos), Interpretativo (impugnar las valoraciones o conclusiones de los datos) e Indemnizatorio (no corresponde a nuestro ordenamiento).

Por otro lado, el *Habeas Data* impuro consiste en solicitar auxilio para recabar información pública que le es negada al afectado. Aquí se trata el *Habeas Data* de Acceso a la Información Pública como tal. Así, el Tribunal constata que, si bien existe una relación de supuestos aplicables al *Habeas Data*, este no debe entenderse como taxativo, sino como enunciativo dependiendo de las situaciones alternas que puedan configurarse con el paso del tiempo.

Es importante indicar que el contenido del *Habeas Data* operará en función a las circunstancias de cada caso concreto, pues el contenido constitucional de un derecho fundamental no puede ser formulado de forma abstracta (Linares, 2019). Siendo esto así, queda en materia del afectado cumplir con los requisitos exigidos por la NCPC (2021), considerando que se analizará tomando las implicancias y alcances particulares aplicables a los derechos protegidos a través del *Habeas Data*.

2.4 Desde la óptica regulatoria peruana: la protección de datos personales

Es oportuno recordar la responsabilidad del Estado en priorizar la seguridad de los datos a ser utilizados por personas naturales, jurídicas e instituciones u organismos públicos y privados por medio de la tecnología, cuando se trate de utilización y procesamiento de cálculos o bases de datos en operaciones de ámbito público o público-privado (Comisión de Ciencia Innovación y Tecnología, 2023, p. 4). Al respecto, el Estado ha reconocido la importancia de contar con una regulación adecuada para el tratamiento de los datos personales, siendo uno de los más resaltantes el Acuerdo Nacional (2017), cuyo acápite 35 advierte que el Estado debe diseñar políticas y regulación para el adecuado resguardo de la seguridad de la información.

Es así que, el 30 de julio de 2011 se publicó la LPDP (2011). Posteriormente, el 22 de marzo de 2013, se complementó con la aprobación del Reglamento (2013), aprobado por el Decreto Supremo No. 003-2013-JUS, el cual entró en vigencia a partir del 8 de mayo de ese mismo año; con la finalidad de garantizar el derecho fundamental a la protección de los datos personales previsto en el artículo 2°, numeral 6 de la Constitución Política del Perú (1993). Seguidamente, la referida LPDP (2011) fue modificada en el 2017 a través del Decreto Legislativo No. 1353 y, con ella, se instauró la Autoridad Nacional de Protección de Datos

Personales (ANPD)[33], que a su vez se encuentra dentro de la Dirección General de Transparencia, Acceso a la Información Pública y Protección de Datos Personales (DGTAIPD) (Decreto Supremo No. 013-2017-JUS, 2017, artículo 70).

Sobre el concepto de datos personales, la LPDP (2011) establece su definición en el numeral 4 del artículo 2º, como "*toda información sobre una persona natural que la identifica o la hace identificable a través de medios que pueden ser razonablemente utilizados".* En ese sentido, el máximo intérprete de nuestra Constitución (Tribunal Constitucional, 2011b, f. 9) señala que por identidad debe entenderse no de modo restrictivo, limitándose a datos que revelen solo señas personalísimas del titular (nombre, sexo, edad, estado civil), sino que la comprensión debe ser amplia al incluir información que revelen aspectos de la identidad relacional, social, económica, política, religiosa, cultura de la persona (desempeño laboral, operaciones comerciales, afiliación política).

Complementando la definición de la LPDP (2011), el numeral 4 del artículo 2 del Reglamento (2013) define a los datos personales como "aquella *información [numérica, alfabética, gráfica, fotográfica, acústica, sobre hábitos personales, o de cualquier otro tipo] concerniente a las personas naturales que las identifica o las hace identificables a través de medios que puedan ser razonablemente utilizados*". Es decir, se considera dato personal a cualquier información que permita identificar o hacer identificable a una persona. Por ejemplo: los nombres y apellidos, la fecha de nacimiento, la dirección del domicilio y la dirección de correo electrónico, el número de teléfono, el número de RUC, el número de la placa del vehículo, la huella digital, el ADN, la voz, una imagen, el número del seguro social, "*likes*" en redes sociales, entre otros, son datos que identifican o son identificables a una persona de manera directa o indirecta conforme se ha analizado en el acápite 2.2 del presente trabajo.

En ese sentido, la aplicación de la referida norma, así como lo que se entiende como datos personales, es aplicable únicamente a las personas naturales, incluyendo a aquellas que cuenten con RUC o personas jurídicas con negocio. En base a dicho artículo, las personas jurídicas quedarían descartadas del ámbito de aplicación (ANPD, 2014), siendo que la razón social, RUC, teléfono, entre otros aspectos no constituyen en datos personales como tal.

[33] A la fecha de presentación de la presente tesis, el 25 de agosto de 2023, la ANPD presentó un Proyecto de Reglamento de la Ley No. 29733, Ley de Protección de Datos Personales, cuyo contenido introduce nuevas figuras en el tratamiento de datos personales, como el derecho a la portabilidad, la instalación de un representante (dentro o fuera del territorio peruano) y la implementación de mecanismos de seguridad como la Evaluación de Impacto del Tratamiento de Datos Personales.

Por otro lado, existe una categoría que cuenta con un especial tratamiento que son los denominados "datos sensibles", los cuales están constituidos por los datos biométricos[34] (huella digital[35], retina, iris), por el cual se derivan aquellos de origen racial y étnico, ingresos económicos, opiniones o convicciones políticas, religión, afiliación sindical y toda información relacionada con la salud y vida sexual. De acuerdo con la ANPD (2017), serán considerados datos sensibles siempre y cuando (i) sean biométricos y (ii) dichos datos, por sí mismos, hagan posible la identificación del titular, es decir, no requiera de alguna otra herramienta para su identificación. Como bien se ha detallado en el punto 2.2 del presente trabajo, se trata de datos personales que únicamente pueden ser objeto de tratamiento con el consentimiento expreso y por escrito del titular de los datos personales.

Respecto al ámbito de la aplicación normativa, en principio el artículo 3 de la LPDP (2011) advierte que corresponde a los datos personales contenidos o destinados a ser contenidos en bancos personales de la administración pública y privada cuyo tratamiento se realiza en el territorio nacional. El Reglamento (2013) extiende y desarrolla tal disposición a través del artículo 5 señalando que el ámbito de aplicación territorial será aplicable en los siguientes supuestos:

- Cuando el tratamiento sea efectuado en un establecimiento ubicado en territorio peruano correspondiente al titular del banco de datos personales o de quien resulte responsable del tratamiento.
- Cuando el tratamiento sea efectuado por el encargado del tratamiento, con independencia de su ubicación, a nombre de un titular de banco de datos personales establecido en territorio peruano o de quien sea el responsable del tratamiento.
- El titular del banco de datos personales o quien resulte responsable del tratamiento no esté establecido en territorio peruano, pero le resulte aplicable la legislación peruana, por disposición contractual o del derecho internacional;

[34] De acuerdo con el ISO/IEC 2382-37:2022(en), la identificación biométrica es el proceso de búsqueda contra una base de datos alojados con el fin de encontrar y obtener uno o varios identificadores de referencias biométricas atribuibles a un solo individuo. Recuperado de: https://www.iso.org/obp/ui/en/#iso:std:iso-iec:2382:-37:ed-3:v1:en:term:37.01.02.

[35] Dentro del Oficio No. 862-2016-JUS/DGPDP, se establece que la huella digital dentro de un reloj marcador de entrada y salida es un dato sensible. Sin embargo, la Autoridad sostiene que se deberá analizar caso por caso, en tanto un reloj por tarjeta no constituiría un dato sensible biométrico, por ejemplo.

- El titular del banco de datos personales o quien resulte responsable no esté establecido en territorio peruano, pero utilice medios situados en dicho territorio, salvo que tales medios se utilicen únicamente con fines de tránsito que no impliquen un tratamiento.

Asimismo, las disposiciones de la LPDP (2011) no serán aplicables, conforme el artículo 3 de la LPDP (2011), en aquellos datos (i) contenidos o destinados a ser contenidos en bancos de datos personales creados por personas naturales para fines exclusivamente relacionados con su vida privada o familiar; ejemplo: una agenda con datos (imágenes, direcciones) de amigos o contactos; y (ii) a los contenidos o destinados a ser contenidos en bancos de datos de administración pública, solo en tanto su tratamiento resulte necesario para el estricto cumplimiento de las competencias asignadas por ley a las respectivas entidades públicas, para la defensa nacional, seguridad pública y para el desarrollo de actividades en materia penal para investigación y represión del delito; ejemplo: bancos de datos que tiene la Policía Nacional del Perú o la Fiscalía por antecedentes penales y/o policiales.

Es importante acotar lo que entendemos por titular del banco de datos personales, el cual, de acuerdo al inciso 17 del artículo 2 de la LPDP (2011), se entiende a toda aquella persona natural, persona jurídica de derecho privado o entidad pública que determina la finalidad y contenido del banco de datos personales, el tratamiento de estos y las medidas de seguridad. Es importante resaltar las funciones y, con ello, la responsabilidad que ostenta un titular del banco de datos personales, pues prácticamente define la totalidad y control del tratamiento de los datos personales.

En esa línea, el Reglamento (2013) introduce un nuevo concepto ligado al titular del banco de datos personales, definiendo como responsable de tratamiento a aquél que decide sobre el tratamiento de datos personales, aun cuando no se encuentren en un banco de datos personales (artículo 2.14).

Por su parte, el numeral 7 del artículo 2 de la LPDP (2011) define al encargado del tratamiento de datos personales como aquella persona natural, persona jurídica de derecho privado o entidad pública, que sola o actuando conjuntamente con otra realiza el tratamiento de los datos personales por encargo -en nombre y por cuenta del titular del banco de datos-, en virtud de una relación jurídica que le vincula con el mismo y delimita el ámbito de su actuación. Incluye

a quien realice el tratamiento sin la existencia de un banco de datos personales. Del mismo modo, debe cumplir con las obligaciones descritas en el artículo 28 de la LPDP (2011).

Sobre el particular, hay que enfatizar que, aunque son conceptos parecidos, constituyen realidades distintas. La ANPD se pronunció sobre la naturaleza y responsabilidades de cada uno de estos actores a través de la Opinión Consultiva No. 034-2021-JUS/DGTAIPD (2021), conforme se detalla a continuación:

Tabla 6.
Diferencias entre titular del banco de datos, encargado de tratamiento y responsable de tratamiento

Titular del banco de datos personales	Encargado de tratamiento	Responsable de tratamiento
• Es la persona (natural o jurídica) que responde por el tratamiento de datos personales almacenados en su banco de datos personales, incluyendo la responsabilidad frente a la finalidad y usos de los mismos. • Ejemplo: un colegio respecto a su banco de datos personales correspondiente a sus estudiantes.	• Se encarga de ejecutar el tratamiento por el cual ha sido encargado bajo dichos límites. • No forma parte de la planilla o del equipo de trabajo del titular del banco de datos personales. • No decide sobre la finalidad del tratamiento de datos personales, sino acatar la finalidad materia del encargo. • No es responsable del tratamiento, salvo que realice tratamientos que no son materia del encargo. • Ejemplo: una empresa contrata los servicios de *cloud* (o alojamiento en la nube).	• Decide sobre el tratamiento de datos personales, aun cuando no se encuentren en un banco de datos personales. • Si no hay encargo de tratamiento, entonces el responsable puede ser igual que el titular del banco de datos. • Ejemplo: una persona natural que decide la difusión o publicación de una imagen que identifica a una persona, una institución educativa, un contador, un médico.

Fuente: Elaboración propia

De otro lado, conforme con el inciso 1 del artículo 2 de la LPDP (2011), se entiende por banco de datos personales a aquel conjunto organizado de datos personales, automatizado o no, independientemente del soporte, sea físico, magnético, digital, óptico u otros que se creen, cualquiera fuere la forma o modalidad de su creación, formación, almacenamiento, organización y acceso. El NCPC (2021) establece, a través de su artículo 53, como banco de datos (o también archivo, registro o base) a todo aquel conjunto de datos organizados de

información personal que sea tratada o procesada en formato físico, electrónico o computarizado, ya sea público o privado, bajo cualquier tipo de modalidad (almacenamiento, formación, acceso y organización). No obstante, un banco de datos será distinto a una base de datos (ANPD, 2015), pues, este último constituye un sistema formado por un conjunto de datos almacenados en discos que permiten el acceso directo a ellos y un conjunto de programas que manipulen ese conjunto de datos. Esta definición solo es una parte de lo que se puede entender por banco de datos personales, por lo que resulta incorrecto señalar que sean sinónimos o cuentan con significados equivalentes.

Otro aspecto relevante es el tratamiento de datos personales**,** entendido como cualquier operación o procedimiento técnico, automatizado o no, que permite la recopilación, registro, organización, almacenamiento, conservación, elaboración, modificación, **extracción**, consulta, utilización, bloqueo, supresión, comunicación por transferencia o por difusión o cualquier otra forma de procesamiento que facilite el acceso, correlación o interconexión de los datos personales (LPDP, 2011, artículo 2.19).

Es importante mencionar que el titular del banco de datos personales debe cumplir con ciertas obligaciones relevantes referidas al tratamiento, como los siguientes:

- No recopilar datos personales por medios fraudulentos, desleales o ilícitos.
- Recopilar datos personales que sean actualizados, necesarios, pertinentes y adecuados, con relación a finalidades determinadas, explícitas y lícitas para las que se hayan obtenido.
- No utilizar los datos personales objeto de tratamiento para finalidades distintas de aquellas que motivaron su recopilación, salvo que medie procedimiento de anonimización o disociación.

Como bien se remarca en el anterior párrafo, la recopilación de datos personales constituye una de las formas de tratamiento de datos personales. Tales formas de tratamiento pueden ser verbales, por escrito, electrónicamente (utilizando formularios online), mediante imágenes con cámaras de videovigilancia, entre otros. En estos casos, quien recoge los datos (el titular del banco de datos o responsable del tratamiento) deberá cumplir con las siguientes obligaciones: a) el deber de informar y b) obtener el consentimiento o la autorización del titular, ello sin

perjuicio de que el titular del dato personal cuenta con el derecho de retirar ese consentimiento en cualquier momento para las finalidades que se haya consentido o solo para alguna de ellas.

Deber de informar

Implica detallar el tratamiento de datos que el responsable del tratamiento o titular del banco de datos realiza, se encuentra regulado en el artículo 18 de la LPDP (2011). Su finalidad es proporcionar al titular de los datos personales de manera detallada, sencilla, expresa, inequívoca y previa a su recopilación[36], de la finalidad y otros aspectos relevantes referidos al tratamiento de los datos personales (Hohfeld, 1968).

Por otro lado, en el caso que el titular del banco de datos establezca vinculación con un encargado de tratamiento de manera posterior al consentimiento, el accionar del encargado queda bajo responsabilidad del titular del banco de datos, debiendo establecer un mecanismo de información personalizado para el titular de los datos personales sobre dicho nuevo encargado de tratamiento.

Este derecho puede considerarse como un principio estricto, en tanto no existe un supuesto de excepción al deber de la información, a diferencia del consentimiento. Por ende, si los datos personales son recogidos en línea, tales aspectos deberán ser debidamente informados y detallados en las políticas de privacidad, las cuales deben ser fácilmente accesibles e identificables dentro del sitio web.

Al respecto, es importante señalar que transmitir los nueve (09) aspectos del artículo 18 exigidos por la LPDP (2011) y el numeral 4 del artículo 12 de su Reglamento (2013) puede resultar complicado en un entorno digital en que cada vez más se valora un diseño sintetizado y apropiado para transmitir la información. De hecho, la lectura de las cláusulas informativas o políticas de privacidad pueden durar horas, resultando dificultosas y no claras para el usuario.

[36] Los nueve aspectos a cumplir conforme con la LPDP (2011) son los siguientes:
- La finalidad para la que sus datos personales serán tratados.
- Quiénes son o pueden ser sus destinatarios.
- La existencia del banco de datos en que se almacenarán.
- La identidad y domicilio de su titular y, de ser el caso, del o de los encargados del tratamiento de sus datos personales.
- El carácter obligatorio o facultativo de sus respuestas al cuestionario que se le proponga, en especial en cuanto a los datos sensibles.
- La transferencia de los datos personales.
- Las consecuencias de proporcionar sus datos personales y de su negativa a hacerlo.
- El tiempo durante el cual se conserven sus datos personales.
- La posibilidad de ejercer los derechos que la ley le concede y los medios previstos para ello.

Esto conlleva a que se incumpla la obligación de presentar la información de forma clara, concisa y sencilla, generando una debilidad en los derechos del titular en recibir la información comprensible.

En ese sentido, la ANPD ha desarrollado una "Guía práctica para la observancia del deber de informar" (2019), cuyo contenido detalla los alcances del deber de información regulado en nuestra LPDP (2011) y Reglamento (2013). Dicho deber se debe materializar (i) de manera previa a la recopilación de los datos personales cuando se obtengan de manera directa y (ii) en el primer contacto con el titular cuando se obtengan de manera indirecta de fuente legítima, como por ejemplo fuentes de acceso públicos (diarios, registros públicos, revistas, guías) o mediante una transferencia.

En adición a ello, la referida Guía (2019) propone que, a fin de evitar una larga extensión de las finalidades de tratamiento, estas se agrupen por categorías, por ejemplo, comunicaciones comerciales, obligaciones legales, entre otras. Del mismo modo, si la relación de los destinatarios de los datos personales es extensa, se puede colocar la categoría del receptor, como por ejemplo instituciones bancarias, servicios de encuestas de calidad; sin perjuicio de que se identifique en otro medio idóneo cada uno de los destinatarios. Además, la ANPD recomienda evitar el uso de citas legales o transcripción de las mismas, así como explicaciones innecesarias del tratamiento de los datos.

Sin embargo, a nuestro juicio, a pesar de contar con la intención de reducir la cantidad de información que se otorga al titular del dato y ofrecer sencillez para una mejor comprensión del mismo, aún no existe una directriz clara y expresa que permita que los titulares de bancos de datos personales o responsables de tratamiento puedan generar propuestas innovadoras referidas a la transmisión de los nueve aspectos del artículo 18 de la LPDP (2011), que serían de utilidad en caso de que el titular del banco de datos personales o responsable de tratamiento ejecute el *Web Scraping*.

Consentimiento

Constituye el eje central de la protección de datos personales (Davara, 2011) y será el titular del banco de datos o el responsable del tratamiento quien cuenta con la carga de la prueba respecto de la validez del consentimiento (Defensoría del Pueblo, 2019). El consentimiento es una manifestación de voluntad que puede ser expresa o tácita según lo contemple cada

ordenamiento jurídico, y que debe ser prestada atendiendo a las circunstancias concretas en las que se solicitan los datos personales (Gacitúa, 2014).

Así, el titular de los datos personales debe conocer, entre otros aspectos, principalmente lo siguiente: la finalidad para la cual sus datos son registrados, el uso de los datos personales y los derechos que asisten a los titulares (Murillo, 1993, pp. 61 y 62). De ahí que la validez del consentimiento para el tratamiento de datos personales se encuentra íntimamente vinculada al cumplimiento de las siguientes condiciones señaladas en el artículo 12 del Reglamento (2013):

- Libre: el artículo 12.1° del Reglamento (2013) indica que el consentimiento se debe haber otorgado sin que medie error, mala fe, violencia o dolo que pueda afectar la manifestación de voluntad del titular de los datos personales. Es decir, debe ser otorgado de manera voluntaria y la persona deberá tener libre elección de denegar o retirar el consentimiento sin sufrir algún perjuicio (Defensoría del Pueblo, 2019), como por ejemplo la no entrega de un beneficio en caso opte por denegar su consentimiento para la finalidad principal. De otro lado, la entrega de beneficios o regalos para obtener el consentimiento no afectarán su validez, con excepción de los casos donde el titular sea un menor de edad, en donde no se considerará libre el consentimiento otorgado mediando obsequios o beneficios. Por otra parte, se debe evitar presentar casillas pre-marcadas dentro de la obtención del consentimiento mediante medios digitales, siendo que no será motivo para negar el servicio o producto que se contrata. Del mismo modo, no se debe solicitar el consentimiento en bloque, es decir para finalidades principales (que no se requiere conforme con el artículo 14 de la LPDP (2011) y para las finalidades adicionales).
- Previo: Referido a la oportunidad en que deberá ser solicitado el consentimiento, esto es, con anterioridad a la recopilación de los datos personales (ANPD, 2014).
- Informado: conforme al artículo 12 numeral 4° del Reglamento (2013), establece que se debe dar a conocer al titular de manera clara, expresa e indubitablemente, con lenguaje sencillo y detallado, los nueve (09) aspectos exigidos por el artículo 18 de la LPDP (2011). Consideramos que, si el consentimiento guarda relación con las transferencias internacionales realizadas, este deberá incluir los posibles riesgos a países terceros en ausencia de contar con el nivel de protección adecuado y/o las garantías pertinentes.

- Expreso e inequívoco: constituye manifestar el consentimiento en condiciones que no admitan dudas de su otorgamiento de manera directa. En otras palabras, se considerará consentimiento expreso a aquel que se manifieste mediante la conducta del titular que evidencie que ha consentido inequívocamente, pues de lo contrario, su conducta hubiere sido otra. La condición de expreso no se limita a la manifestación verbal o escrita tradicional, pues dentro del entorno digital, se considera expresa la manifestación de "hacer clic", "cliquear" o "pinchar", "dar un toque", "touch" o "pad" u otros similares. En ese sentido, el consentimiento escrito puede otorgarse mediante firma electrónica, siempre que quede grabada de manera que pueda ser impresa o leída, o que por medio de cualquier otro mecanismo o procedimiento establecido se permita identificar al titular y recabar su consentimiento, a través del texto escrito. Entonces, el consentimiento expreso se refiere a que debe haberse dado a través de un acto positivo como una casilla de selección electrónica que el titular puede marcar en línea o la firma en un formulario. Por tanto, el silencio, las casillas premarcadas o la inacción, no deben constituir el consentimiento.

Sin embargo, dado que ningún derecho es absoluto, la LPDP (2011) regula, en su artículo 14° y sus normas modificatorias, las trece (13) excepciones respecto de la obligación de solicitar el consentimiento al titular de los datos personales para su tratamiento[37]. Resulta importante

[37] Ley de Protección de Datos Personales, Ley No. 29733
Artículo 14. Limitaciones al consentimiento para el tratamiento de datos personales
No se requiere el consentimiento del titular de datos personales, para los efectos de su tratamiento, en los siguientes casos:

- **A)** Los datos personales son recopilados o transferidos por entidades públicas para el cumplimiento de sus funciones (artículo 14 numeral 1). Ejemplo: La ONP recopila datos de sus aportantes para cumplir con el pago de una pensión de jubilación, conforme a ley.
- **B)** Los datos personales se encuentran contenidos en fuentes accesibles al público. Ejemplo: El portal de transparencia estándar de una entidad pública contiene datos (nombre y DNI) de las personas que mantienen reuniones de trabajo con diversos funcionarios. Lo que se entiende por fuentes de acceso público se encuentran regulados en el artículo 17 del Reglamento de la LPDP (2011).
- **C)** Se hace tratamiento de datos personales relativos a la solvencia patrimonial y de crédito. Ejemplo: Las entidades financieras tienen acceso a datos personales relativos a la solvencia patrimonial o de crédito de sus clientes, para analizar la prestación de sus servicios.
- **D)** Cuando medie norma para la promoción de la competencia en los mercados regulados emitida en ejercicio de la función normativa por los organismos reguladores a que se refiere la Ley No. 27332, Ley Marco de los Organismos Reguladores de la Inversión Privada en los Servicios Públicos, o la que haga sus veces, siempre que la información brindada no sea utilizada en perjuicio de la privacidad del usuario.
- **E)** Los datos personales son necesarios para la preparación, celebración y ejecución de un contrato en el que el titular de los datos es parte. Ejemplo: Para comprar un inmueble se firmará un contrato entre la constructora y el particular, el cual debe consignar los datos personales que permitan identificar plenamente al comprador.
- **F)** Se realiza el tratamiento de datos personales de salud, en circunstancias de emergencia, para prevenir, diagnosticar y tratar al titular, en centros de salud y por profesionales de la salud, observando el secreto profesional. Aunado a ello, la ANPDP sostiene que, será factible el tratamiento de los datos relacionados con la salud sin consentimiento del titular, cuando (i) el tratamiento deba realizarse por razones de salud pública, calificada por el Ministerio de Salud; y (ii) para la realización de estudios epidemiológicos o análogos, debiendo aplicarse procedimientos de disociación adecuados.
- **G)** Se hace tratamiento de datos de los miembros de entidades sin fines de lucro, cuya finalidad sea política, religiosa o sindical, siempre que se circunscriba a sus miembros, sus actividades y que no sean transferidos sin consentimiento.

advertir que, si bien no es vital la obligación de contar con el consentimiento del titular en tales situaciones, de igual forma se debe cumplir con el deber de información y las demás obligaciones y principios regulados en la LPDP (2011) y su Reglamento (2013).

En el caso de datos sensibles, el consentimiento deberá ser obtenido por escrito a través de una firma manuscrita, firma digital o cualquier otro mecanismo de autenticación que garantice la voluntad inequívoca del titular conforme con el artículo 14º del Reglamento (2013). Aunado a ello, y en línea con el inciso 6 del artículo 13 de la LPDP (2011), el tratamiento de datos sensibles puede efectuarse cuando la ley lo autorice siempre y cuando medie motivos de interés público.

Por otro lado, el titular de los datos personales tiene el derecho a exigir que sus datos se utilicen de manera adecuada por parte del titular del banco de datos, o el encargado de tratamiento de sus datos; de acuerdo a los derechos que se detallan a continuación:

- Acceso: Toda persona tiene derecho a obtener la información que sobre sí mismo sea objeto de tratamiento ya sea en bancos de datos de administración pública o privada, en específico sobre la forma en que sus datos fueron recopilados, las razones que motivaron su recopilación y a solicitud de quién se realizó la recopilación, además de las transferencias realizadas y planeadas a futuro.

Ejemplo: La recopilación de datos de trabajadores adscritos a un sindicato, que permita la realización de sus actividades propias.

H) Se hubiera aplicado un procedimiento de anonimización o disociación. Ejemplo: La Autoridad Nacional de Protección de Datos Personales publica los informes y consultas que emite en el ejercicio de sus funciones, para lo cual tacha los nombres y domicilios de las personas que intervinieron en el procedimiento.

I) El tratamiento de datos personales resulta necesario para salvaguardar intereses propios del titular. Ejemplo: Ante la desaparición de una persona, las autoridades policiales tratan y publican sus datos personales para facilitar su ubicación.

J) Los datos personales son tratados para prevenir el lavado de activos, activos, financiamiento del terrorismo u otros delitos, por mandato legal. Ejemplo: Investigaciones iniciadas por el Ministerio Público por la presunta comisión del delito de lavado de activos, sin requerir consentimiento del investigado/a.

K) En caso de grupos económicos conformados por las empresas que califican como sujetos obligados a informar a la Unidad de Inteligencia Financiera (UIF- Perú) deben compartirse información entre sí de sus clientes para prevenir el lavado de activos y financiamiento del terrorismo y delitos conexos, así como otros de cumplimiento regulatorio, estableciendo las salvaguardas adecuadas sobre la confidencialidad y uso de la información intercambiada. Ejemplo: Las empresas del sistema financiero y sistema de seguros, la bolsa de valores, las cooperativas de ahorro y crédito y otros sujetos obligados comparten información de sus clientes, con fines preventivos, según la Ley 27693.

L) El tratamiento de datos se realiza en virtud del derecho fundamental de libertad de información. Ejemplo: Medios de comunicación publican noticias de interés, relacionadas a la coyuntura política o social brindando datos personales de personas involucradas en hechos noticiosos.

M) Otros que deriven del ejercicio de competencias expresamente establecidas por Ley.

- Rectificación (Actualización, Inclusión): Es el derecho que posee el titular para que modifiquen sus datos que resulten ser parcial o totalmente inexactos, incompletos, erróneos o falsos.
- Información: El titular tiene derecho a ser informado respecto del tratamiento de sus datos, incluyendo la finalidad, los destinatarios, el banco en el que se almacenarán, el tiempo de conservación y lo relacionado con el tratamiento.
- Cancelación (Supresión): Es el derecho del titular para solicitar la supresión o cancelación de sus datos personales de un banco de datos personales cuando éstos hayan dejado de ser necesarios o pertinentes considerando la finalidad para la cual fueron extraídos; hubiere vencido el plazo establecido previamente para su tratamiento; haya revocado su consentimiento y casos adicionales en los que no están siendo tratados conforme a la LPDP (2011) y su Reglamento (2013).
- Oposición: Toda persona tiene la posibilidad de oponerse, por un motivo legítimo y fundado, referido a una situación personal concreta, a figurar en un banco de datos o al tratamiento de sus datos personales, siempre que por una ley no se disponga lo contrario y sin que hubiera brindado su consentimiento.

Estos derechos son también denominados “Derechos ARCO” y se ejercen de forma personal y gratuita a través de solicitudes dirigidas al titular del banco de datos o al encargado del tratamiento.

Del mismo modo, la LPDP (2011) y su Reglamento (2013) establecen los siguientes derechos que ostenta el titular del dato:

- Derecho a impedir el suministro: respecto a sus datos personales conforme con el artículo 21 de la LPDP (2011). No aplica para la relación del titular del banco de datos personales y el encargado de tratamiento.
- Derecho al tratamiento objetivo: derecho a no verse sometido a una decisión con efectos jurídicos obre el titular del dato que le afecte de forma significativa sustentado únicamente en un tratamiento de sus datos personales basado en determinados aspectos de su personalidad o conducta, salvo excepciones reguladas por el artículo 23 de la LPDP (2011).
- Derecho a la tutela: en caso el titular o encargado de datos no cumpla con el debido y adecuado tratamiento como tal, el titular del dato tiene derecho a recurrir por la vía de

la ANPD o el Poder Judicial mediante la acción de *Habeas Data*. Este derecho se encuentra regulado en el artículo 24 de la LPDP (2011).

- Derecho a ser indemnizado: goza de tal derecho el titular del dato personal a causa del incumplimiento por parte del titular o encargado de tratamiento conforme con las disposiciones de la LPDP (2011). Regulado en el artículo 25 de la LPDP (2011).

Finalmente, la LPDP (2011) establece ciertas obligaciones que deben cumplir los titulares de los bancos de datos personales, junto con el responsable y encargado de tratamiento, al momento de recabar, registrar, almacenar, tratar, transferir, difundir y utilizar datos personales. En caso no se respeten los derechos ARCO, se prevé las siguientes dos vías para resguardar su derecho, conforme con el artículo 24 de la LPDP (2011):

A) La vía judicial, a través del proceso de *Hábeas Data* (artículo 200.3 de la Constitución), donde el titular puede conocer, actualizar, incluir, suprimir o rectificar la información referida a su persona que se encuentre almacenada o registrada en forma manual, mecánica o informática, en archivos, bancos de datos o registros de entidades públicas o privadas que brinden servicios o acceso a terceros. Asimismo, puede suprimir o impedir que se suministren datos de carácter sensible que afecten su intimidad. Este proceso fue debidamente explicado en el acápite 2.3.2.1 del presente trabajo.
B) La vía administrativa, por medio de un procedimiento trilateral de tutela ante la ANPD conforme el artículo 24 de la LPDP (2011), en cuyo caso el titular del dato requiere el acceso, rectificación, corrección u oposición (derechos ARCO) a cualquier tratamiento referido a su persona que se encuentre almacenado o registrado en forma manual, mecánica o informática en archivos, bancos de datos o registros de entidades públicas o privadas que brinden servicios o acceso a terceros o a falta de pronunciamiento o negativa a dicha solicitud. Los casos pueden ser porque el titular, quien no es un personaje de interés público, cuenta con una publicación en el Internet relacionado a un evento pasado negativo (DGTAIPD, 2019), o porque el titular estuvo involucrado con un sujeto con problemas judiciales y/o policiales (DGTAIPD, 2018).

   Otro tipo de procedimiento administrativo que tiene a cargo la ANPD son los procedimientos administrativos sancionadores, los cuales buscan proteger los derechos sobre los datos personales desincentivando tratamientos proscritos por ley (Luna, 2021,

p. 259). Tal potestad sancionadora finaliza con la imposición de multas tipificadas en cada una de las infracciones del artículo 132 del Reglamento (2013), las cuales son calificadas como: (i) "Leves" (0.5 a 5 UIT)[38], (ii) "Graves" (> 5 a 50 UIT)[39] y (iii) "Muy Graves" (>50 a 100 UIT)[40]. Las multas serán analizadas bajo la Metodología para el Cálculo de las Multas en materia de Protección de Datos Personales (Ministerio de Justicia y Derechos Humanos, 2020). Durante el 2023, la ANPD recaudó más de S/ 7,6 millones a raíz de la fiscalización de 336 entidades (Ministerio de Justicia y Derechos Humanos del Perú, 2023), en cuyo caso se espera que dicho número siga incrementándose a fin de resguardar la protección de los datos personales.

CAPÍTULO III: MEDIDAS REGULATORIAS EXTRANJERAS

3.1 Análisis Comparado

Una vez analizado nuestro esquema nacional peruano, es imprescindible remitirnos a la regulación extranjera que, de forma directa o latente, ha constituido una fuente inspiracional para la elaboración de la regulación peruana en materia de protección de datos personales. Es por ello que, en el presente capítulo, exploraremos las distintas normativas de protección de datos personales, las cuales inciden en la extracción y uso de los datos personales en línea dentro de los países pertenecientes a la Unión Europea, Estados Unidos, y los principales países de Latinoamérica.

En específico, la regulación extranjera ha dado un paso adelante y se ha pronunciado sobre la técnica del *Web Scraping* o extracción masiva de datos personales por medios automatizados mediante jurisprudencia y normativa. Es tal su impacto que dicha práctica ha sido identificada por distintas Autoridades de Protección de Datos Personales como Reino Unido, Colombia, México y Argentina, como un riesgo para el adecuado tratamiento de datos personales y que, por tanto, merece el establecimiento de medidas técnicas y organizativas que permitan mitigar tal riesgo.

[38] Considerar que, mediante Decreto Supremo No. 309-2023-EF, 1 UIT equivale a S/ 5,150 (Cinco Mil Ciento Cincuenta y 00/100 Soles). Dicha infracción puede llegar a representar hasta S/ 25,750 (Veinticinco mil setecientos cincuenta con 00/100 Soles).
[39] Dicha infracción puede llegar a representar hasta S/ 257,500 (Doscientos cincuenta y siete mil quinientos con 00/100 Soles).
[40] Dicha infracción puede llegar a representar hasta S/ 515,000 (Quinientos quince mil con 00/100 Soles).

Un importante ejemplo de ello reincide en la Declaración conjunta sobre *Web Scraping* y protección de datos (Information Commissioner's Office [ICO], 2023) [41], cuyas disposiciones señalan, entre otros aspectos, que **las empresas de redes sociales y los operadores de sitios web que alojan datos personales** (de acceso público) **deben proteger los datos de los titulares contra el *Web Scraping* ilegal**. Asimismo, advierte que la disponibilidad de los datos en el internet no implica carta abierta para darle tratamiento a los datos personales, ya que estos se encuentran protegidos por la normativa de protección de datos aplicable.

Si bien la referida Declaración constituye un avance respecto a la protección de datos personales frente a la técnica del *Web Scraping*, queda pendiente una mayor visibilidad y pronunciamiento de los demás países, así como el entendimiento de su alto impacto negativo frente a la privacidad de los usuarios. Por ello, se han identificado a los principales países del extranjero que han dado cuenta de la técnica del *Web Scraping* y se detallará el razonamiento utilizado en el marco de la existencia de dicha práctica.

3.1.1 Europa

Europa contiene una particularidad especial al ser uno de los continentes que ha desarrollado regulaciones características al acopio o extracción de datos personales y excepciones al consentimiento. Ello iniciando desde la Declaración Universal de los Derechos Humanos, adoptado el 10 de diciembre de 1948 en Paris, para luego emitirse el Convenio Europeo de Derechos Humanos, adoptado por el Consejo de Europa el 4 de noviembre de 1950[42].

Luego de ello, el 28 de enero de 1981 se emitió el Convenio No. 108, para la protección de las personas con respecto al tratamiento automatizado de datos de carácter personal (Consejo de Europa, 1981), y que es considerado el primer texto jurídico vinculante con vocación universal

[41] Suscrita por la Autoridad de Datos de Australia, Canadá, Reino Unido, China, Suiza, Noruega, Nueva Zelada, Colombia, Jersey, Marruecos, Argentina y México.

[42] Dentro del primer capítulo, el derecho a la protección de datos se encuentra recogido implícitamente en el artículo 8, dedicado a la vida privada y familiar: "1. Toda persona tiene derecho al respeto de su vida privada y familiar, de su domicilio y de su correspondencia. 2. No podrá haber injerencia de la autoridad pública en el ejercicio de este derecho sino en tanto en cuanto esta injerencia esté prevista por la ley y constituya una medida que, en una sociedad democrática, sea necesaria para la seguridad nacional, la seguridad pública, el bienestar económico del país, la defensa del orden y la prevención de las infracciones penales, la protección de la salud o de la moral, o la protección de los derechos y las libertades de los demás." En definitiva, se puede apreciar el carácter limitado de los derechos fundamentales, pues para adoptar ciertas decisiones que causen la injerencia en la vida privada, se debe superar el juicio de ponderación.

en el ámbito de la protección de datos[43]. Suscrito por Alemania, Francia, Dinamarca, Austria y Luxemburgo, el Convenio contaba con la finalidad de garantizar, en el territorio de cada Parte, a cualquier persona natural sea cual fuera su nacionalidad o su residencia, el respeto de sus derechos y libertades fundamentales y, en particular, de su derecho a la vida privada respecto al tratamiento automatizado de sus datos de carácter personal, entendiéndose por "*datos de carácter personal*" cualquier información relativa a una persona física identificada o identificable (persona concernida), y por "*tratamiento automatizado*" a las operaciones de registro de datos, aplicación a dichos datos de operaciones lógicas aritméticas, su modificación, borrado, extracción o difusión, y operaciones efectuadas en su totalidad o en parte con ayuda de procedimientos automatizados. El otro objetivo (subordinado) del Convenio (1981) es "*la libre circulación de la información* ", es decir, liberalizar la transferencia de datos personales[44].

Seguidamente, se emitió la Directiva 95/46/CE del Parlamento Europeo y del Consejo (1995), cuyo contenido regulaba a los datos sometidos a tratamiento –total o parcialmente- automatizados (base de datos informática de clientes, por ejemplo), así como al tratamiento de los datos personales contenidos en un fichero no automatizado o que vayan a figurar en él (ficheros en papel tradicionales).

Posteriormente, se creó la Carta de los Derechos Fundamentales de la Unión Europea (2000), cuyo contenido reconoce el respeto a la vida privada y la protección de datos personales como derechos fundamentales estrechamente relacionados, pero independientes[45]. Así, también se cuenta con el Tratado de Lisboa (2009, artículo 16-B) que ratifica lo indicado en la citada Carta (2000).

Luego de ello, la Directiva 95/46/CE (1995) quedó derogada por el Reglamento Europeo de Protección de Datos Personales (2016) (en adelante, "RGPD*"*). Tras un período transitorio de

[43] Hasta ahora, lo han firmado y ratificado 46 de los 47 Estados Miembros del Consejo. El último Estado en ratificarlo, durante el 2013 fue Uruguay. Para mayor información: http://conventions.coe.int/Treaty/Commun/ChercheSig.asp?NT=108&CM=8&DF=09/07/2013&CL=ENG

[44] El Convenio No. 108 (1981) también plantea una serie de principios, siendo la principal la "calidad de los datos" (art. 5), desde el cual se organiza el resto de las normas y principios. Los principios que vienen de la calidad de los datos son: el principio de veracidad, el de seguridad y el principio de finalidad. Por su parte, los derechos de acceso, rectificación y cancelación constituirían "garantías complementarias" a los principios señalados, a favor de los afectados.

[45] Este documento reconoce por una parte el derecho al respeto a la vida privada y familiar (artículo 7), y por otro, el derecho a la protección de datos personales (artículo 8). Es decir, consagra ambos derechos de forma independiente y autónoma. Tal artículo se basa en el artículo 286 del Tratado Constitutivo de la Comunidad Europea y en la Directiva 95/46/CE del Parlamento Europeo y del Consejo relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos (DO L 281 de 23.11.1995, p. 31), así como en el artículo 8 del CEDH y en el Convenio del Consejo de Europa para la protección de las personas con respecto al tratamiento automatizado de datos de carácter personal, ratificado por todos los Estados miembros.

dos años, se convirtió en norma directamente aplicable en todos los Estados miembros de la Unión Europea el 25 de mayo de 2018.

A raíz del RGPD (2016), se creó el Comité Europeo de Protección de Datos - CEPD (*European Data Protection Board*), un organismo independiente que vela por el cumplimiento adecuado del referido cuerpo jurídico.

La aplicación del RGPD (2016), a simple vista, podría implicar únicamente a todas las organizaciones establecidas en la UE. Sin embargo, para una organización no establecida en la UE puede aplicarse el RGPD (2016) si trata datos personales de interesados en la UE cuando las actividades de tratamiento se relacionan "con la oferta de bienes o servicios" (RGPD, artículo 3, apartado 2, letra a) a esos interesados en la UE o "con el seguimiento de su comportamiento" (RGPD, artículo 3, apartado 2, letra b).

En cuanto a la definición de dato de carácter personal que adopta el RGPD (2016) es "toda información sobre una persona física identificada o identificable (<el interesado>)". En esa línea, se considera persona física identificable a toda persona cuya identidad pueda determinarse de manera directa o indirecta con un identificador, como, por ejemplo, los de geolocalización, número de teléfono, placa de automóvil, entre otros. Estos identificadores deberán contar con uno o varios elementos propios de la identidad que incluyen los siguientes aspectos: física, fisiológica, genética, psíquica, económica, cultural o social de dicha persona. Es decir, constituyen datos personales los datos de identificación directa (p.e. nombres y apellidos) y los datos seudonimizados o información de identificación no directa (p.e. sustitución por códigos o técnicas de tokenización).

En la actualidad, el RGPD (2016) profundiza y abarca el concepto de dato de manera más amplia, pues con las Tecnologías de la Información y Comunicaciones (TIC) y las herramientas tecnológicas y/o digitales como redes sociales, *cloud computing*, aplicaciones, entre otros; las personas pueden ser identificadas de diversas maneras dentro del entorno digital para otorgar seguridad jurídica a los tratamientos masivos de datos personales.

Con referencia al **tratamiento de los datos personales**, el RGPD (2016) permite realizar el tratamiento de los mismos, incluyendo la extracción, siempre y cuando se cuente con una **base legal**, es decir, una razón permisible especificada en la ley. De ese modo, el RGPD (2016)

describe seis bases jurídicas (RGPD, art. 6) que pueden justificar la recolección y/o extracción de datos personales:

(1) El **interesado ha dado su consentimiento** para el procesamiento de sus datos personales para uno o más fines específicos;
(2) El procesamiento es **necesario para la ejecución de un contrato** al que el interesado es parte o para adoptar medidas a petición del interesado sujeto antes de celebrar un contrato;
(3) El procesamiento es **necesario para el cumplimiento de una obligación legal** a la que el responsable del tratamiento está sujeto;
(4) El procesamiento es **necesario para proteger los intereses vitales** de los datos sujeto o de otra persona natural[46];
(5) El procesamiento es indispensable **para la realización de una tarea con el fin de salvaguardar el interés público o por las funciones que ejerce una autoridad oficial** conferida al controlador;
(6) El procesamiento es **necesario para los fines de los intereses legítimos perseguidos por el responsable del tratamiento o por un tercero.** salvo que tales intereses sean anulados por los intereses o derechos fundamentales y libertades del interesado que requieren la protección de datos personales, en particular cuando el interesado sea un niño. Esto será aplicable siempre que sea por satisfacer intereses legítimos de terceros. Por lo que deberá realizarse un test de ponderación aplicable al caso concreto y documentarlo para poder justificar dicho tratamiento sobre la base de esta base legitimadora. En esa misma línea, y a pesar de que no se encuentre de forma expresa, se debe documentar e identificar claramente la legitimación sobre la que se fundamentan los tratamientos, de acuerdo con el RGPD (2016) y el principio general de "responsabilidad proactiva" (RGPD, art. 5)[47]. Es importante destacar que las disposiciones del RGPD (2016) no se aplican a los datos anonimizados conforme con el Recital 26 del RGPD (2016).

[46] En la práctica se utiliza el presente inciso para realizar el tratamiento de los datos personales sin el consentimiento.
[47] Este principio establece la necesidad de que los responsables y encargados de tratamiento demuestren que han tomado las medidas técnicas y organizativas necesarias para garantizar que los tratamientos sean conforme al RGPD (2016) mediante documentación o actividades como: análisis de riesgo, medidas de seguridad, registro de actividades de tratamiento, protección de datos desde el diseño y por defecto, notificación de violaciones de seguridad de los datos, evaluación de impacto sobre la protección de datos, delegado de protección de datos, entre otros equivalentes.

Se aprecia que, entre las situaciones que permiten el tratamiento lícito de datos sin consentimiento, la satisfacción del interés legítimo es la excepción con mayor aplicación práctica en el RGPD (2016), pues se suele recurrir en calidad de cajón de sastre para dejar de obtener el consentimiento del interesado en determinadas situaciones en las que resulta de difícil o imposible obtención. Al respecto, dentro del Considerando 47 del RGPD (2016), se establece la regla de la expectativa legítima o razonable para la aplicación de dicho inciso bajo el interés legítimo[48].

Al respecto, dentro de los Considerandos 47 al 49 del RGPD (2016), se citan ejemplos para la aplicación de tal supuesto. Así, el Considerando 47 establece la posibilidad de aplicar el interés legítimo cuando exista una relación pertinente y apropiada entre el interesado y el responsable y se espere de forma razonable que dicho tratamiento tendrá lugar (por ejemplo, una relación con un cliente o trabajador). En el Considerando 48 se establece el tratamiento de datos con fines de marketing directo, prevención del fraude, transmisiones de datos dentro de un grupo empresarial; y en el Considerando 49 se refiere a las transmisiones de datos para garantizar la seguridad de las redes.

Además de dichas bases legitimadoras para tratar los datos personales, **la categoría de datos sensibles requiere de otra base adicional para su tratamiento**. Para el RGPD (2016), los datos sensibles incluyen aquellos datos que revelen el origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas, o la afiliación sindical, y el tratamiento de datos genéticos[49], datos biométricos dirigidos a identificar de manera unívoca a una persona física, datos relativos a la salud o datos relativos a la vida sexual o la orientación sexual de una persona física[50].

Según el artículo 9 del RGPD (2016), existen distintas excepciones sobre el tratamiento de datos sensibles, siendo los siguientes supuestos:

[48] Dicha regla es establecida por el Grupo de trabajo 29 en Opinión No. 6 de 2014.

[49] De acuerdo con el inciso 13 del artículo 4 del RGPD (2016), se entiende como datos genéticos a aquellos relativos a las características genéticas heredadas o adquiridas de una persona física que proporcionen una información única sobre la fisiología o la salud de esa persona, obtenidos en particular del análisis de una muestra biológica de tal persona.

[50] Artículo 9 del RGPD

1. Quedan prohibidos el tratamiento de datos personales que revelen el origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas, o la afiliación sindical, y el tratamiento de datos genéticos, datos biométricos dirigidos a identificar de manera unívoca a una persona física, datos relativos a la salud o datos relativos a la vida sexual o la orientación sexual de una persona física.

1. Se cuenta con el consentimiento[51] expreso del interesado
2. Fines laborales y de seguridad social
3. Salvaguardar intereses vitales del interesado o de otra persona natural
4. Fines caritativos
5. El interesado manifestado públicamente dichos aatos personales
6. Reclamaciones legales o judiciales
7. Interés público esencial
8. Medina preventiva o laboral, usos sanitarios y diagnósticos médicos
9. Fines de interés público en la salud pública; y
10. Fines de interés público para la investigación o fines estadísticos

En la práctica, la principal diferencia entre el tratamiento de datos personales y datos sensibles recae en que este último no puede realizarse por la causal de interés legítimo (RGPD, art. 6), sino únicamente por los supuestos detallados en el artículo 9 del RGPD (2016).

Tal limitación es importante, en tanto los intereses legítimos son una justificación comúnmente usada para procesar datos sin consentimiento, como por ejemplo ejecutar la técnica del *Web Scraping.* Así, en caso de no ocupar alguna de las excepciones detalladas en el artículo 9, se deberá obtener el consentimiento expreso del titular del dato.

De igual forma, el RGPD (2016) ha traído una novedad respecto a las disposiciones ligadas a la realidad tecnológica, ya que advierte que en caso se realice un tratamiento de datos particular al utilizar nuevas tecnologías y que implique un alto riesgo para los derechos de las personas, entonces el responsable de tratamiento deberá realizar una Evaluación de Impacto relativa a la protección de datos (en adelante, "DPIA") a fin de identificar los riesgos y acciones frente a estos. Ello se aplicará específicamente cuando se traten datos sensibles (RGPD, 2016, art. 35).

Sin perjuicio de los alcances de la obtención del consentimiento, se debe cumplir con el deber de información detallado en el artículo 13 del RGPD (2016), siempre y cuando el consentimiento sea obtenido directamente por el interesado:

[51] De acuerdo con el inciso 11 del artículo 4 del RGPD, el consentimiento debe ser una "indicación libre, específica, informada e inequívoca de los datos voluntad del sujeto por el cual o ella, mediante una declaración o una clara afirmación acción, significa aceptar el procesamiento de datos personales que le conciernen".

- La identidad y los datos de contacto del responsable y, en su caso, de su representante;
- Los datos de contacto del delegado de protección de datos, en su caso; de su representante;
- Los fines del tratamiento a que se destinan los datos personales y la base jurídica del tratamiento;
- Cuando el tratamiento se base en el artículo 6, apartado 1, letra f), los intereses legítimos del responsable o de un tercero;
- Los destinatarios o las categorías de destinatarios de los datos personales, en su caso;
- La intención del responsable de transferir datos personales a un tercer país u organización internacional y la existencia o ausencia de una decisión de adecuación de la Comisión, o, en el caso de las transferencias indicadas en los artículos 46 o 47 o el artículo 49, apartado 1, párrafo segundo, referencia a las garantías adecuadas o apropiadas y a los medios para obtener una copia de estas o al hecho de que se hayan prestado;
- Plazo de conservación o los criterios que permiten delimitar el plazo;
- Medios para el ejercicio de los derechos de los titulares, como acceso, rectificación, supresión, limitación, oposición y portabilidad de los datos;
- Derecho a presentar un reclamo ante una autoridad de control;
- Si la comunicación de datos personales es un requisito legal o contractual y si el interesado está obligado a facilitar los datos personales y está informado de las consecuencias en caso de no se otorgue la facilitación del tratamiento; y
- La existencia de decisiones automatizadas, incluida la elaboración de perfiles[52].

En el supuesto de que no se haya obtenido el consentimiento directamente por el interesado, entonces entramos en el ámbito de aplicación del artículo 14 del RGPD (2016), cuyo contenido advierte que deberán informarse los aspectos del artículo 13 del RGPD (2016) en un momento posterior: (i) dentro de un plazo razonable después de la obtención de los datos que no exceda de un mes; (ii) si los datos personales están destinados a la primera comunicación con el interesado, entonces que no exceda de una semana; y (iii) si está prevista la divulgación a un tercero, entonces que se realice cuando los datos personales se divulguen por primera vez.

[52] Conforme con el inciso 4) del artículo 4 del RGPD (2016), se entiende por elaboración de perfiles a toda forma de tratamiento automatizado de datos personales consistente en utilizar datos personales para evaluar determinados aspectos personales de una persona física, en particular para analizar o predecir aspectos relativos al rendimiento profesional, situación económica, salud, preferencias personales, intereses, fiabilidad, comportamiento, ubicación o movimientos de dicha persona física.

Sin embargo, contamos con la excepción del artículo 14.5 b) del RGPD (2016)[53], el cual advierte que no será necesario proporcionar los aspectos del artículo 13 del RGPD (2016) si resulta imposible o constituye un esfuerzo desproporcionado.

En caso alguno de que los responsables o encargados del tratamiento incumplieran las disposiciones del RGPD (2016), estos últimos facultan a las autoridades supervisoras a imponer multas de hasta el 4% de la facturación mundial anual, es decir, hasta veinte millones de euros, sin perjuicio de las medidas correctivas aplicables. Tales multas deberán ser efectivas, proporcionadas y disuasorias.

Ahora bien, cabe preguntarse **si a raíz de las bases legitimadoras de tratamiento de datos personales y sensibles, entonces es posible la ejecución del *Web Scraping***. Al respecto, el hecho de que un dato personal se encuentre publicado o sea accesible por internet, así como que sea indexable, no implica que, únicamente por este hecho, sea legítima su recolección y extracción (tratamiento), pues como se ha advertido previamente, se requiere de una base legitimadora conforme con el artículo 6 del RGPD (2016) y, del mismo modo, del artículo 13 del RGPD (2016) aplicable a los datos sensibles.

En materia de *Web Scraping*, los supuestos aplicables al tratamiento de datos personales al amparo del RGPD (2016) sería o bien que se cuente (i) con el consentimiento del titular, o bien (ii) se realice una tarea de interés público, o bien (iii) se cuente con un tener un interés legítimo para tratar los datos y que dicho tratamiento sea necesario para lograr tal interés.

De los supuestos mencionados, el más factible a utilizar (y de hecho el más usado) es el caso (iii) donde el responsable del tratamiento alega un interés legítimo. Actualmente, el RGPD (2016) no define expresamente lo que se debe entender por interés legítimo; sin embargo, como se ha señalado anteriormente en los Recitales 47 al 49 del RGPD (2016), se advierte una posible respuesta a dicha incógnita: siempre que no prevalezcan los intereses o los derechos y libertades

[53] Artículo 14. Información que deberá facilitarse cuando los datos personales no se hayan obtenido del interesado
5. Las disposiciones de los apartados 1 a 4 no serán aplicables cuando y en la medida en que:
b) la comunicación de dicha información resulte imposible o suponga un esfuerzo desproporcionado, en particular para el tratamiento con fines de archivo en interés público, fines de investigación científica o histórica o fines estadísticos, a reserva de las condiciones y garantías indicadas en el artículo 89, apartado 1, o en la medida en que la obligación mencionada en el apartado 1 del presente artículo pueda imposibilitar u obstaculizar gravemente el logro de los objetivos de tal tratamiento. En tales casos, el responsable adoptará medidas adecuadas para proteger los derechos, libertades e intereses legítimos del interesado, inclusive haciendo pública la información.

fundamentales del interesado, considerando las expectativas razonables de los mismos basadas en su relación con el responsable del tratamiento.

De ese modo, se debe evaluar caso por caso si es que el interesado pudo esperar razonablemente, en el momento y contexto de la extracción de los datos personales mediante el *Web Scraping* que el tratamiento haya tenido lugar con tal finalidad.

Por ello, se deberá evaluar (i) el interés legítimo del responsable de tratamiento (el que ejecuta el *Web Scraping*); (ii) la necesidad del tratamiento para lograr ese interés; (iii) analizar que se protejan sus derechos y libertades fundamentales de los titulares de datos; y (iv) las expectativas razonables del titular. Este último elemento resulta controversial, en tanto pueden darse supuestos subjetivos de lo que entendemos por expectativas razonables de un titular.

Al respecto, consideramos que las expectativas razonables dependen de la funcionalidad y/o naturaleza del *website*. Por ejemplo, si dentro de Facebook o LinkedIn coloco visible mi teléfono celular o correo electrónico, entonces estaría dejando abierta la posibilidad de que me identifiquen y contacten por dichos medios. Además, la Política de Privacidad de la *website* puede dilucidar el cómo se tratarán los datos personales y, con ello, contar con ciertas expectativas en cuanto a su tratamiento.

Además de ello, se deberá cumplir con el deber de informar si se realiza tratamiento de datos personales en razón a un legítimo interés, salvo que constituya un esfuerzo desproporcional [54] conforme con el artículo 14.5 b) del RGPD (2016). Tal "esfuerzo desproporcional" será analizado caso por caso[55] y debidamente sustentado para su aplicación.

Asimismo, se requerirá de una DPIA cuando se realice tratamiento a gran escala mediante el *Web Scraping*, pues se trata de un tipo de "*procesamiento de datos invisible*". El DPIA es un

[54] Para la Autoridad de Datos Personales de Polonia, los altos costos para cumplir con el deber de información no constituyen un esfuerzo desproporcional. Para más información, ver https://iapp.org/news/a/polish-court-overturns-dpas-first-RGPD (2016)-fine/.

[55] La Autoridad de Datos Personales de Polonia (Polish Data Protection Authority) impuso una multa ascendente a 220,000 euros a un ciudadano sueco por haber realizado extracción de datos personales de fuentes oficiales del territorio que involucraba a 7,5 millones de titulares. El demandado alegó que había colocado la información requerida por el RGPD (2016) dentro de su sitio web, ya que constituía un alto costo contactar a una gran parte de los titulares, siendo que, a su parecer, se encontraba frente a la excepción del artículo 14.5 b). Para mayor información, puede acceder al siguiente enlace: https://uodo.gov.pl/pl/324/787.

análisis minucioso realizado por el responsable del tratamiento, quien evalúa el daño potencial de los interesados y las medidas de respaldo que tomará a fin de mitigar dicho riesgo.

En cuanto al tratamiento de datos sensibles, debido a que no se encuentra el supuesto de interés legítimo, entonces únicamente se podría realizar *Web Scraping* conforme los supuestos previamente citados de acuerdo al artículo 9 del RGPD (2016), entre ellos, el que se haya obtenido el consentimiento del titular del dato. Sin embargo, dado que en la práctica no se realiza el *Web Scraping* conforme los supuestos habilitadores del RGPD (2016) ni se obtiene el consentimiento como tal, entonces no se recomienda realizar *Web Scraping* sobre datos sensibles, pues, de lo contrario, se estaría infringiendo con el RGPD (2016) y bajo apercibimiento de sanción por parte de autoridad de datos personales del territorio.

Con respecto a casos especiales de la normativa europea, se cuenta con la Directiva 2019/790 (2019), en el que dentro del artículo 3 habilita a los organismos de investigación e instituciones responsables del patrimonio cultural a realizar reproducciones y extracciones a poder realizar (...) minería de texto y datos de obras u otras prestaciones a las que tengan acceso lícito. Sin embargo, en el artículo 4 establece limitaciones o excepciones al regular la posibilidad de restringir la extracción de las bases de datos y su contenido.

- Existe una excepción obligatoria para las universidades y otros organismos de investigación, respecto del derecho exclusivo de reproducción y del derecho de prohibir la extracción de una base de datos (artículo 4.1). El artículo 4 numeral 2 no establece un plazo de conservación si es el tiempo necesario para cumplir la minería de textos y datos.
- Para empresas privadas, esta excepción debe aplicarse cuando los titulares de derechos no se hayan reservado adecuadamente los derechos de reproducciones y extracciones para minería de datos. (artículo 4.3).

Sobre el segundo caso, el titular de tales bases de datos podrá restringir su uso realizando una reserva expresa de su derecho a través de la publicación de tal restricción, como por ejemplo un *pop up o robots.txt* que informe que la extracción de datos (o *Web Scraping*) está prohibido[56].

[56] Únicamente se encontraría habilitado el *Web Scraping* en casos de investigación científica, conforme con el artículo 3 de la propia Directiva.

De lo señalado previamente, el *Web Scraping* no se encuentra prohibido expresamente por el RGPD (2016); sin embargo, en la práctica resulta complicada su aplicación en tanto se deben considerar las bases legitimadoras para el tratamiento de los datos personales y sensibles, siendo que, en la mayoría de los supuestos revisados, no se podrá ejecutar el tratamiento de datos vía *Web Scraping* de forma legítima.

A) España

La legislación española en materia de protección de datos personales ha sido considerada como fuente inspiracional y ejemplar por distintos países de Latinoamérica, quien a su vez esta última tuvo como referentes a otros países de la región, como Alemania. El origen de la regulación de protección de datos personales fue con la Constitución (1978), en la que se incluyó el derecho a la protección de datos a través del artículo 18.4. de la Carta Magna.

A través de la Sentencia 254/1993 (1993) se estableció el derecho de los ciudadanos a conocer los datos personales que le conciernen y que se hallan registrados en archivos administrativos informatizados. Posteriormente, mediante la Sentencia 292/2000 (2000), el Tribunal Constitucional español declaró que el derecho a la protección de datos personales es un derecho fundamental autónomo, independiente y diferenciado por los demás derechos regulados en el artículo 18.4 de la Constitución (1993)[57]. Tal derecho consiste en que la persona cuenta con poder sobre el control de sus mismos datos personales, evitando su uso ilícito y lesivo contra la dignidad del titular.

Además del reconocimiento constitucional, es importante resaltar que las fuentes inspiracionales para la legislación española provienen del Convenio No.108 del Consejo de Europa (1981); y de las facultades que se otorgan a las personas reconocidas en la práctica norteamericana de protección frente a los informes de solvencia patrimonial desde los sesenta del siglo XX, en la que presupone que el consentimiento es la base de la protección de datos personales y se necesita del consentimiento inequívoco de los afectados, debidamente

[57] Como hemos señalado anteriormente, será con la STC 292/2000, donde el Tribunal Constitucional Español despejará las ambigüedades y establecerá rotundamente que el derecho a la protección de datos es un derecho autónomo e independiente. De esta forma, pasa de ser considerado un «instituto de garantía de los derechos a la intimidad y el honor y del pleno disfrute de los restantes derechos de los ciudadanos», a constituirse, también, en «un derecho fundamental».

informados, o con autorización legal explícita para confirmar su licitud (García Gonzáles, cita a Murillo y Piñar, 1990).

Es así que, en base al Convenio No. 108 del Consejo de Europa (1981), se creó la Agencia Española de Protección de Datos ("AEPD"), la cual se caracteriza por ser un ente de derecho público con personalidad jurídica propia y plena capacidad pública y privada. Su propósito constituye en ser un órgano especializado y facultado en la investigación, atención de las reclamaciones de los afectados, la resolución de conflictos, así como ejercer la potestad inspectora y sancionadora, entre otras, en materia de protección de datos personales (Castillo, 1994, p. 361). A la fecha, la AEPD cumple un rol activo al no solo cumplir con lo señalado previamente, sino también educar sobre el valor de la protección de datos por parte del individuo mediante la publicación de documentos orientativos y charlas al público en general.

Asimismo, el 29 de octubre de 1992 se aprueba la Ley Orgánica de Regulación del Tratamiento Automatizado de los Datos de Carácter Personal (en adelante, "LORTAD") (1992), cuyo objetivo consistía en regular el uso de la informática cuando estaban implicados datos personales y buscaba además garantizar el derecho a la privacidad. Al regular solo el tratamiento automatizado, los tratamientos documentales y audiovisuales no se incluían dentro del ámbito de su aplicación. En consecuencia, solo el infractor que cometía la infracción a través de un soporte informático podía ser sujeto a sanción al amparo de la regulación de la LORTAD.

La LORTAD tuvo influencia de la entonces Directiva 95/46/CE (1995) con vigencia en todo el territorio europeo, la cual delimitaba[58] qué se entendía por datos personales, señalando en su artículo 2 letra a) que es "*toda información sobre una persona física identificada o identificable"* (el interesado). Del concepto cabe distinguir la exclusión respecto de los datos de personas jurídicas, pues se trata de proteger un derecho fundamental.

Posteriormente, se emitió la Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal (en adelante, "LOPD"), cuyo ámbito de aplicación abarcaba tanto para el tratamiento de datos automatizados como no automatizados (soporte físico). Del mismo modo, se mantuvo la definición de dato personal, manteniéndose la referencia exclusivamente a la

[58] Esta Directiva es la fuente legislativa directa de todas las legislaciones protectoras de datos de los países miembros de la Unión Europea, en cuyo caso se han adecuado a la referida directiva.

información relativa a las personas físicas[59]. Al respecto, la doctrina española analizó los principales aspectos compuestos por la definición de dato personal (Grupo de Trabajo del Artículo 29, 2007):

1. Cualquier información: Este término implica un concepto amplio de protección de datos.
    a. Desde el punto de vista de contenido, no es una lista taxativa, sino que se pueden incorporar conceptos a medida que evoluciona la tecnología y la sociedad en conjunto (Gil, 2016), como los datos personales en el entorno digital. Por ejemplo: la geolocalización, la dirección IP, entre otros.
    b. Desde el punto de vista de la naturaleza de la información, el contenido incluye tanto información objetiva como las evaluaciones subjetivas (Gil, 2016). Las evaluaciones subjetivas son atribuciones que se le imponen a un sujeto determinado de acuerdo con patrones de comportamiento. En algunos sectores como la banca ("prestatario moroso"), seguros ("no se espera que Juan muera pronto") o el laboral ("trabajador productivo"), tales evaluaciones son bastante comunes. Entonces, no es necesario que la información sea veraz, pues el interesado o titular tiene el derecho de acceder a dicha información y rectificarla a través de los medios adecuados para ejercer sus derechos.
    c. Desde el formato o soporte, se permite la aparición de datos de forma alfabética, fotográfica, sonora o cualquier otro. Esto es fundamental considerando que no es necesario que la información se encuentre almacenada en una base de datos o en un fichero estructurado, sino que es válido que sea contenida en un texto en un documento electrónico. Por ejemplo: las instrucciones grabadas en un celular.

2. Persona identificada o identificable: La segunda característica se divide en dos.

[59] Entiéndase como dato personal a la dirección de correo electrónico (Informe Jurídico de la AEPD. Ref. de entrada 1999-9910. Recuperado en https://www.aepd.es/es/documento/1999-9910.pdf), la dirección IP (Informe Jurídico de la AEPD. Ref. de entrada 327-2003. Recuperado en https://www.aepd.es/es/documento/2003-0327.pdf), el número de teléfono (Informe Jurídico de la AEPD. Ref. de entrada 285-2006. Recuperado en https://www.aepd.es/es/documento/2006-0285.pdf), la matrícula de un vehículo (Informe Jurídico de la AEPD. Ref. de entrada 184-2006. Recuperado en https://www.aepd.es/es/documento/2006-0184.pdf), la voz (Informe Jurídico de la AEPD. Ref. de entrada 1999-9905. Recuperado en https://www.aepd.es/es/documento/1999-9905.pdf), la imagen (Informe Jurídico de la AEPD. Ref. de entrada 2001-9908. Recuperado en https://www.aepd.es/es/documento/2001-9908.pdf) huella digital (Informe Jurídico de la AEPD. Ref. de entrada 1999-9903Recuperado en https://www.aepd.es/es/documento/1999-9903.pdf).

a. Persona identificada: Lo es cuando la información disponible indica que pertenece directamente a alguien sin tener que realizar ninguna investigación posterior.
b. Persona identificable: Lo es cuando, aunque no haya sido identificada, pueda serlo en conjunto con otros datos, siempre que la identificación no requiera. de actividades o plazos desproporcionados[60]. Este enfoque permite un alcance dinámico y extensivo de la aplicación de la regulación española, así como también cualquier dato que se asocie a datos personales se convierte en dato personal. Por ejemplo, si una persona no identificada es propietaria de un gato, ello no es un dato personal; sin embargo, cuando este hecho se vincula a datos que pueden identificar a la persona, como su teléfono de celular, entonces este hecho se convierte en un dato personal. Del mismo modo, la identidad de una persona es posible con una cantidad de datos no identificables. Por ejemplo, combinando datos que indican que una persona no identificada es dueña de un gato siamés, información sobre las mejores veterinarias del radio, programas de cuidado para gato siamés, entre otros, entonces se podría llegar a la identificación del dueño.

De lo revisado, la doctrina española atribuye a que la finalidad del tratamiento, es decir, el provecho que busca obtener el responsable del tratamiento, así como el costo de la identificación, son importantes para evaluar la identificación de un individuo. Por ello, no basta la mera e hipotética posibilidad de singularizar a un individuo (Gil, 2016).

Luego de la LOPD, se emitió el Real Decreto 1720/2007, de 21 de diciembre, considerado como el Reglamento de desarrollo de la LOPD (en adelante, "RDLOPD") (2007). Un aspecto novedoso fue el consentimiento de los menores de edad, pues estableció la licitud del consentimiento de los menores de edad desde los 14 años[61]. Asimismo, el RDLOPD establece

[60] Considerando 26 de la Directiva 95/46/CE del Parlamento Europeo y del Consejo, de 24 de octubre de 1995, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos; y artículo 5 del Real Decreto 1720/2007, del 21 de diciembre, por el que se aprueba el Reglamento de Desarrollo de la Ley Orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal.

[61] Esta cuestión suscitó un debate en la doctrina jurídica por dos motivos principalmente. En primer lugar, por el carácter reglamentario de esta norma ya que lo más oportuno hubiese sido en formato de ley. Pese a que la Ley Orgánica 1/9996, de 15 de enero sobre Protección Jurídica del Menor, ex art. 4, ya hace lo suyo al amparar la intimidad personal y familiar. En segundo lugar, al determinar la edad mínima en 14 años se torna la problemática sobre las verdaderas condiciones de madurez del menor para tomar decisiones sobre el tratamiento automatizado de sus datos. MARTÍNEZ MARTÍNEZ, R. (2008). "El real decreto 1720/2007, de 21 de diciembre, por el que se aprueba el reglamento de desarrollo de la ley orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal. aspectos clave", Revista Jurídica de Castilla y León, nº 16, pp. 274-275.

una diferenciación de funciones entre el responsable y encargado del tratamiento de datos personales, en tanto el responsable del tratamiento tenía la obligación adicional de validar que el encargado del tratamiento reúna las garantías autosuficientes dentro de sus servicios en materia de protección de datos personales, quedando la posibilidad de subcontratar servicios, con aprobación del responsable de tratamiento, una vez cumplida la relación contractual (p.e. labores de conservación o destrucción). Además, el encargado de tratamiento asume las responsabilidades y sanciones del responsable.

Posteriormente, la LOPD fue derogada y sustituida por la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales (en adelante, "LOPDGDD") (2018), cuyo contenido adapta el RGPD (2016) al ordenamiento jurídico español, así como también incluye mejoras significativas en el marco del entorno digital. Como parte de tales iniciativas mejoras aplicables al ámbito digital, es la aceptación de nuevos derechos digitales, como el acceso universal a internet, neutralidad de la red, educación digital y el derecho al olvido.

Un aspecto significativo de esta norma es la inclusión del concepto de información previa en la recogida de datos personales de los interesados en el artículo 11 del LOPDGDD (2018). De esta forma, en caso se obtengan datos personales de un interesado, el responsable deberá suministrar información primaria sobre su tratamiento (primera capa), tales como: la identidad del responsable del tratamiento, la finalidad del tratamiento, la legitimidad, destinatarios, la viabilidad de ejercer sus derechos (acceso, rectificación, supresión, oposición, limitación del tratamiento y portabilidad) e información adicional mediante un enlace para consultar con mayor detalle el tratamiento de los datos. En calidad de segunda capa, se desplegará con mayor énfasis, explicación y mención sobre el tratamiento de los datos del interesado, como los plazos o criterios de conservación de los datos, categorías de los datos tratados o los datos de contacto del Delegado de Protección de Datos (AEPD, 2018).

Con referencia a lo que podemos considerar como fuentes de acceso al público, la AEPD ha determinado que las páginas webs no son fuentes accesibles al público, y que, por ello, deberá obtenerse el consentimiento debido para su tratamiento (AEPD, 2008). Asimismo, ello se avala en la Sentencia del Tribunal de Justicia de la Unión Europea (2011) y la Sentencia de la Audiencia Nacional (2012).

En cuanto al tratamiento de datos personales mediante la técnica del *Web Scraping*, se cuenta con un caso emblemático de Equifax Ibérica, S.L. (en adelante, "Equifax") (2019). El caso incide en que Equifax, una agencia multinacional de informes de crédito al consumo, extrajo información sobre deudas de las personas (datos personales sensibles) mediante fuentes públicas. Tal información publicada en la web tenía como finalidad brindar su servicio de acceso a su Fichero de Reclamaciones Judiciales y Organismos Públicos, cuyo contenido albergaba obligaciones adeudadas con la administración pública por parte de personas naturales y jurídicas.

Tras un análisis de las acciones cometidas por Equifax, la AEPD advierte que el infractor buscaba obtener un beneficio económico utilizando la información sobre la solvencia de los interesados; en consecuencia, dispuso que Equifax infringió: (i) el principio de limitación de la finalidad (art. 5.1.b); (ii) el principio de licitud (art. 6.1); (iii) el principio de exactitud (art. 5.1.d); (iv) el principio de minimización de datos (art. 5.1.c) y (v) la obligación de informar al interesado en caso no se hayan obtenido los datos directamente (art. 14). Por ello, la AEPD impuso una multa de aproximadamente un millón de euros y le ordenó eliminar los datos ya recopilados, así como dejar de recopilar datos por ser contrario al RGPD (2016).

Una vez revisado el esquema regulatorio, se concluye que España no presenta muchos casos relacionados a la técnica de *Web Scraping*; no obstante, la AEPD sigue a la vanguardia en cuanto a las nuevas tecnologías y medidas a considerar para salvaguardar tal derecho fundamental y constitucionalmente protegido[62], cumpliéndose con los requisitos de **proporcionalidad y finalidad**.

B) Alemania

Alemania es considerada un país pionero en materia legislativa de protección de datos, siendo que el 07 de octubre de 1970 se publicó la *Ley del Land de Hesse (Datenschutzgezetz*) (1970). Tal norma otorgaba protección a las personas cuyos datos personales estaban sujetos de tratamiento por parte de las entidades públicas alemanas o personas jurídicas de derecho público. Para ello, se implementó al Comisario Parlamentario de Protección de Datos (*Datenschtzauftragrer*) con matices similares a las de un Oficial de Cumplimiento

[62] Por tanto, si bien la Constitución no le impone límites específicos, ni remite a los poderes públicos para su determinación, los límites al ejercicio del derecho a la protección de datos han de encontrarse en los restantes derechos fundamentales y bienes jurídicos constitucionalmente protegidos.

(*Ombudsman)*, considerado un magistrado público - o un tercero - que se encargaba de la observancia de cumplimiento de la referida norma.

Años después, se emitió la *Bundesdatenschutzgesetz* o la Ley Federal de Protección de Datos (1977), cuyo contenido establecía disposiciones favorables para la protección de datos en el espacio privado como público, como la obligatoriedad de la implementación de medidas técnicas y organizativas por parte del responsable de tratamiento; y la incorporación del Delegado de Protección de Datos en toda corporación que contaba por lo menos con cinco personas permanentemente.

Seguidamente, Alemania aprobó distintas normativas sectoriales, como *Teledienstedatenschutzgesetz* (1996), que regula el tratamiento de datos personales en las telecomunicaciones y servicios de información, o el *Bundesdatenschutzgesetz für den Bereich des Gesundheitswesens* (1995) que regula el tratamiento de datos sensibles en el ámbito sanitario. Ello fue hasta que, el 25 de mayo de 2018, Alemania adoptó al RGPD (2016) mediante la aprobación de la Ley Federal Alemana de Protección de Datos (*Bundesdatenschutzgesetz* - "BDSG") (2018). Tal norma regula las numerosas cláusulas de apertura del RGPD (2016) que permiten a los Estados miembros especificar o incluso restringir los requisitos de procesamiento de datos conforme con el RGPD (2016). Además del BDSG (2018) y el RGPD (2016), Alemania emitió la Ley de Telecomunicaciones-Telemedios-Protección de Datos (T*elekommunikation-Telemedien-Datenschutzgesetz)* (2021), cuyo instrumento jurídico ha servido en calidad de esclarecedor en interpretaciones del BDSG (2018) y RGPD (2016) en entornos digitales. Tal norma sostiene (Sección 12) que todo proveedor de servicios podrá extraer y utilizar datos personales para otros fines siempre y cuando el destinatario del servicio haya dado su aprobación, además de excepciones reguladas por ley.

Alemania cuenta con autoridades en cada uno de los dieciséis estados federales alemanes (*Länder*) y, al mismo tiempo, con el Comisionado Federal Alemán para la Protección de Datos y la Libertad de Información (*Bundesbeauftragter für Datenschutz und Informationsfreiheit*)[63], quien es la autoridad supervisora de todos los organismos públicos federales y representa a Alemania en el Comité Europeo de Protección de Datos.

[63] Para mayor información puede visitar la página web institucional: https://www.bfdi.bund.de/DE/Home/home_node.html.

Con referencia a la definición de dato personal y dato sensible, la jurisdicción alemana mantiene las mismas disposiciones del RGPD (2016). Sobre ello, es importante resaltar el análisis de de los casos *Scarlet Extended* (2011) *y Patrick Breyer v. Bundesrepublik Deutschland*, en donde la *Bundesgerichtshof* (Corte Federal de Justicia de Alemania) formaliza un criterio respecto a qué debemos entender por "*medios razonables*" para que configure un dato personal. En tales pronunciamientos, Alemania adopta el enfoque en que solo deben considerarse los medios que "*puedan utilizarse razonablemente*", es decir, sostiene que la determinación de si algo es razonable depende de los obstáculos técnicos y jurídicos, además del esfuerzo (por ejemplo, en cuanto a tiempo, dinero y/o mano de obra) se requiera. Si el esfuerzo es desproporcionado (por ejemplo, realizar una ardua investigación para dar a conocer la dirección IP y solicitarla vía judicial), entonces los medios no deben considerarse razonables y, por tanto, no estamos frente a un dato personal.

En cuanto al procesamiento de datos personales, se mantiene las bases legitimadoras señaladas en el RGPD (2016). Sin embargo, dentro del tratamiento de datos sensibles, el BDSG (2018) deroga el artículo 9 (1) del RGPD (2016) y establece nuevos supuestos para el tratamiento de datos sensibles[64] por los privados (Sección 22 del BDSG):

- Cuando el tratamiento es imperioso para el ejercicio de derechos derivados del derecho a la Seguridad Social y a la protección social y para el cumplimiento de las obligaciones correspondientes;
- Cuando el tratamiento es necesario para fines de medicina preventiva, para la evaluación de la capacidad laboral del trabajador, el diagnóstico médico, la prestación de asistencia sanitaria o social o tratamiento o la gestión de sistemas y servicios de asistencia sanitaria o social o en virtud del contrato del interesado con un profesional sanitario y si estos datos son tratados por profesionales sanitarios u otras personas sujetas a la obligación de secreto profesional o bajo su supervisión;
- Cuando el tratamiento es esencial por razones de interés público en el ámbito de la salud pública, incluyendo la protección frente a amenazas transfronterizas agresivas para la salud o la garantía de un alto nivel de calidad y seguridad de asistencia sanitaria y de

[64] Es decir, datos personales que revelen el origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas, o la afiliación sindical, y el tratamiento de datos genéticos, datos biométricos dirigidos a identificar de manera unívoca a una persona física, datos relativos a la salud o datos relativos a la vida sexual o las orientaciones sexuales de una persona física.

medicamentos o productos sanitarios; además de las medidas contempladas en el apartado 2, deberán cumplirse, en particular, las disposiciones de Derecho laboral y penal destinadas a garantizar el secreto profesional; o bien

- Cuando el tratamiento sea urgentemente relevante por razones de interés público importante.

Asimismo, para el tratamiento de datos sensibles, se adoptarán medidas apropiadas y específicas[65] para salvaguardar los intereses del interesado, en cuyo caso se considerarán los costes de aplicación, el estado de la técnica, el alcance, el contexto y los fines del tratamiento, así como los riesgos de probabilidad y gravedad variables para los derechos y libertades de las personas físicas que plantea el tratamiento. Es importante acotar que el BDSG (2018) no deroga el artículo 9 (2) del RGPD (2016) referida a las excepciones de tratamiento de datos sensibles conforme se detalló anteriormente, por lo que, por ejemplo, se podría realizar el tratamiento de datos sensibles si se cuenta con el consentimiento del titular del dato.

Del mismo modo, será posible tratar datos personales para una finalidad distinta en el que se recogieron los datos (Sección 24 del BDSG) si (i) el tratamiento es necesario para prevenir amenazas a la seguridad pública o del Estado o para perseguir delitos; o (ii) el tratamiento es necesario dentro de un procedimiento judicial. Lo anterior será posible siempre y cuando el titular del dato tenga un interés superior en que no se traten los datos.

En caso de incumplimiento, el responsable de datos personales puede afrontar multas de hasta cincuenta mil euros (Sección 43 del BDSG), además de afrontar penas privativas de la libertad de hasta tres años (Sección 42 del BDSG).

El *Web Scraping* se encuentra regulada conforme con el RGPD (2016) y el BDSG (2018), en cuyo caso debe cumplir con todas las disposiciones de tratamiento de datos personales y sensibles para ejecutar el *Web Scraping* dentro del territorio.

A la fecha, Alemania no cuenta con casos emblemáticos relacionados a la extracción de datos personales a través del *Web Scraping*. Sin embargo, el Tribunal Federal de Justicia Alemán

[65] Tales medidas pueden ser las medidas técnicas organizativas del RGPD (2016), medidas para aumentar la concienciación del personal implicado en las operaciones de tratamiento, la seudonimización de los datos personales, cifrado de datos personales, entre otros.

(*Bundesgerichtshof)* (2014) señaló que el *Web Scraping* se encontraba permitido, incluso si se extraían datos personales que se encuentran disponibles públicamente, siempre y cuando la finalidad no sea para un fin ilegal ni viole el derecho de los interesados.

Si trasladáramos dicho caso a la actual normativa vigente, es decir el RGPD (2016) y BDSG (2018), es evidente que el *Web Scraping* se encontraría supeditado a las bases legitimadoras establecidos en ambos cuerpos normativos, considerando que todo dato personal y/o sensible disponible en la web, por el simple hecho de ser accesible, no corresponde legítimo su tratamiento.

C) Italia

A diferencia de los demás países europeos, Italia tardó en reconocer el derecho constitucional a la protección de datos personales, en tanto no fue hasta la Ley No. 675 y No. 676, ambas del 31 de diciembre de 1996, en donde se reguló la protección de datos personales. Lo resaltante de tales cuerpos normativos corresponden a que, a pesar de haberse emitido en un periodo donde el internet no era comúnmente utilizado a nivel global, se regulaban supuestos de transferencia de datos personales al extranjero[66], incluyendo la obligación de que el país receptor cuente con niveles adecuados de protección para salvaguardar el derecho de las personas. Asimismo, se consideraban sanciones administrativas, así como penas privativas de la libertad de hasta dos años[67].

Actualmente, Italia está dentro del régimen del RGPD (2016), aprobado a través del Decreto Legislativo 101/2018, siendo que a su vez se modificaron disposiciones del Decreto Legislativo 196/2003 (2003) (también denominado "Código de Privacidad"). Italia cuenta con una autoridad de control denominado "G*arante per la protezione dei dati personali*"[68] ("Garante"),

[66] Ley de 31 de diciembre de 1996, n. 675
Art. 28. (Transferencia de datos personales al extranjero).
1. La transferencia, incluso temporal, fuera del territorio nacional, por cualquier forma o medio, de los datos personales objeto de tratamiento deberá ser notificada previamente al Garante, si se dirige a un país no perteneciente a la Unión Europea o se refiere a alguno de los datos a que se refieren los artículos 22 y 24.
[67] Ley de 31 de diciembre de 1996, n. 675
Art. 34. (Notificación omitida o infiel).
1. El que, estando obligado a ello, no realice las notificaciones exigidas en los artículos 7 y 28, o indique en ellas datos incompletos o falsos, será reprimido con prisión de tres meses a dos años. Si el hecho se refiere a la notificación prevista en el párrafo 1 del artículo 16, la pena será de prisión de hasta un año.
[68] Para mayor información puede visitar la página web institucional: https://www.garanteprivacy.it/temi/internet-e-nuove-tecnologie/dating-online.

quien está compuesto por un Consejo (cuatro personas elegidas por autoridades del Estado) y una Oficina.

En cuestión del tratamiento de datos, las disposiciones se basan por el RGPD (2016) y el Código de Privacidad (2003). Este último sostiene que el tratamiento de datos especiales (por ejemplo, datos sensibles) necesarios para la ejecución de un fin con interés público está permitido si la legislación europea o nacional prevé el tratamiento o por un acto administrativo. Para ello, se deberá identificar los motivos de interés público por las que se realiza el tratamiento, los tipos de datos, las operaciones a realizarse y las medidas apropiadas y específicas de protección de los interesados.

Bajo este esquema, el tratamiento de datos genéticos, biométricos o relativos a la salud deberán cumplir requisitos adicionales que serán identificados por el Garante mediante medidas específicas (artículo 2- sexies, 2003). Aunado a ello, se deberá considerar las directrices, recomendaciones y mejores prácticas en materia de seguridad y tratamiento de datos personales publicadas por el CEPD. De otro lado, cuando sea necesario difundir o transferir datos personales a otros sujetos por razones de interés público, será necesario notificarlo al Garante al menos 10 días antes de dicho tratamiento (artículo 2- ter, 2003).

Bajo la extracción de datos personales mediante el *Web Scraping* dentro del territorio italiano, se aplican las medidas dispuestas en el RGPD (2016), es decir, deberá contarse con las bases legitimadoras para el tratamiento de los datos personales y sensibles. Al respecto, el Garante se ha pronunciado sobre controversias relacionadas al *Web Scraping*, siendo la más reciente en el año 2023 mediante la Decisión No. 201 (2023), en donde terminó imponiendo una multa ascendente a €60,000 euros y una orden correctiva al propietario del sitio web ”www.trovanumeri.com“ por difundir números de teléfono en una guía telefónica, obtenidos a partir del *Web Scraping*. La medida correctiva incluía la prohibición de extraer, almacenar y publicar datos personales para la creación y difusión en línea de una guía telefónica a través de su página web.

Por otro lado, se cuenta con la Orden judicial contra Clearview AI (2022), pues se determinó que la referida empresa habría extraído ilegalmente datos personales y sensibles (biométricos) en territorio italiano mediante la técnica del *Web Scraping*. La finalidad de tal extracción se direccionaba para el entrenamiento de su sistema de LLM (*Large Language Models*), en cuyo

caso se incluía el tratamiento de imágenes de redes sociales (por ejemplo, Twitter o Facebook), blogs y sitios web donde están presentes fotografías de acceso público, pero también de vídeos disponibles en línea (por ejemplo, en Youtube). A partir de dichos datos sensibles, utilizaban técnicas biométricas para extraer las características identificativas de cada una de ellas y, posteriormente, transformadas en "representaciones vectoriales" para indexar y alimentar el sistema LLM.

Dado que se trataba de las imágenes extraídas mediante *Web Scraping* para brindar un servicio de búsqueda biométrica altamente calificado (coincidencia facial), el Garante no dudó en imponer una multa ascendente a €20.000.000 euros por no gozar legitimidad en el tratamiento de datos conforme con el RGPD (2016).

Finalmente, el Garante anunció el lanzamiento de una consulta pública sobre *Web Scraping* dirigida a todo tipo de usuarios web (públicos o privados) establecidos en Italia o que ofrecen sus servicios en Italia en los que se divulgan públicamente datos personales (2023). El motivo de la investigación incide en consolidar una serie de recomendaciones, lineamientos o medidas técnicas para evitar la extracción masiva de datos personales mediante el *Web Scraping* con fines de entrenamiento de algoritmo de IA.

Es así que, el pasado 20 de mayo de 2024, la autoridad italiana (Garante) emitió una nota orientativa y acciones de prevención contra la extracción masiva de datos personales vía *Web Scraping* (GPDP, 2024). Advirtió que la evaluación de licitud del tratamiento vía *Web Scraping* deberá darse caso por caso, y las medidas preventivas a adoptar dependerán de cada titular de banco de datos o responsable de tratamiento para proteger la información alojada en su plataforma digital o página web. Si bien estas medidas son incapaces de impedir el Web Scraping en su totalidad o al 100%, deberán implementarse siguiendo el principio de responsabilidad proactiva o *accountability.*

Dichas citadas medidas pueden ser la (i) creación de “áreas reservadas” en las que se puede acceder mediante previo registro para que se pueda tratar los datos personales, (ii) dentro de los Términos y Condiciones señalar expresamente que se encuentra prohibido el *Web Scraping*, (iii) monitorear el tráfico de la red imponiendo, de ser el caso, limitaciones de acceso si existieran actividades sospechosas e (iv) intervención de bots que mitiguen la extracción

indiscriminada de datos personales, como la inserción de CAPTCHAs en la plataforma digital o página web.

D) Reino Unido

La historia de la regulación de protección de datos personales en Reino Unido inició en 1984 a través del *Data Protection Act*, la primera norma sobre la materia y que reconoció, como puntos resaltantes, los derechos de información, acceso, rectificación y cancelación de los titulares, así como también la creación de órganos jurisdiccionales especiales que revisaban asuntos sobre la materia (*Information Rights Tribunal*).

Años más tarde, la citada norma fue derogada por el *Data Protection Act* de 1998 y que su contenido regulatorio respondía al contenido de la Directiva 95/46/CE vigente en ese entonces. Como se ha explicado, dicha norma solo aplicaba a personas naturales, descartando la posibilidad sobre la regulación de personas jurídicas. En ese mismo año, el Reino Unido aprobó el *Human Rights Act* de 1998, el cual reconocía en un ámbito constitucional el derecho a la privacidad[69], constituyendo un hito en protección a la privacidad para la población del Reino Unido.

Posteriormente, en el 2016 ocurre el fenómeno político del *Brexit*[70], lo cual trajo como consecuencia que el Reino Unido deje de pertenecer a la Unión Europea y, con ello, la normativa aplicable a dicho conjunto de países. A pesar de lo anterior, el gobierno británico optó por mantener el RGPD (2016) adoptado a ciertos cambios técnicos que permitan dilucidar su plano nacional, creando así el "*RGPD del Reino Unido*". Así, se cuenta con la Ley de Protección de Datos (Data Protection Act, 2018) que se encuentra vigente y complementa al RGPD del Reino Unido, regulando aspectos específicos, como por ejemplo las bases legitimadoras para el tratamiento de datos especiales.

[69] Human Rights Act 1998
Article 8 - Right to respect for private and family life
1. Everyone has the right to respect for his private and family life, his home and his correspondence.
2. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

[70] El 23 de junio de 2016 se realizó un referéndum para conocer si es que la población del Reino Unido deseaba mantenerse en la Unión Europea. El resultado fue el de abandonar la EU y se procedió con la desvinculación de la misma, que se extendió hasta el 31 de enero de 2020.

En otro orden de cosas, la aplicación del RGPD del Reino Unido dependerá si (i) la organización está establecida en el territorio, (ii) trata datos personales que se encuentren en el Reino Unido cuando el tratamiento está relacionado "*con la oferta de bienes o servicios*" o (iii) "*con la supervisión de su comportamiento*" en la medida en que su comportamiento tenga lugar en el territorio.

Un aspecto interesante, y que no corresponde a los países de la UE, es que los responsables de tratamiento deben pagar tasas anuales para ejecutar dicho tratamiento, el cual se cuantifica en función al número de empleados y del volumen de ventas anual, así como también si el responsable del tratamiento es una entidad pública o una ONG[71]. En caso de no cumplir con la citada obligación, entonces el responsable del tratamiento tendrá que afrontar una multa de hasta 4.350 GBP por parte del ICO, entidad encargada de la protección de datos personales en el territorio inglés.

En cuanto a los principios básicos de tratamiento, estos corresponden a que deben tratarse de forma lícita, leal y transparente ("principio de licitud, lealtad y transparencia"); con fines específicos y expresos ("principio de limitación de la finalidad"); deben ser adecuados y pertinentes con la finalidad o finalidades ("principio de minimización de datos"); precisos y renovados (principio de exactitud); conservados por el tiempo necesario ("principio de limitación del almacenamiento"); y tratados con una seguridad adecuada ("principio de integridad y confidencialidad"). Tales principios deberán ser cumplidos a cabalidad y demostrar su cumplimiento fehacientemente conforme con el principio de rendición de cuentas (*accountability principle*), por lo que las figuras de auditorías y políticas que avalen una adecuada gobernanza de datos son imprescindibles para todo tipo de tratamiento de datos.

Las bases legales de tratamiento son iguales a las estipuladas en el RGPD de la Unión Europea, en donde se cuenta con los mismos supuestos legales habilitadores para el tratamiento, en los que se incluye el consentimiento del titular de datos y la tutela de los intereses legítimos del responsable del tratamiento o de un tercero siempre que no transgreda los derechos y libertades fundamentales del interesado. Asimismo, el tratamiento de datos especiales se encuentra restringido, salvo supuestos expresos en la norma, y que deberá considerar los supuestos habilitadores para su tratamiento.

[71] La tasa máxima a cancelar por parte de los responsables de tratamiento con gran volumen de ventas y trabajadores asciende a 2.900 GBP.

En cuanto a la aplicación del deber de información, se aplican las disposiciones del artículo 13 y 14 del RGPD de la Unión Europea; sin embargo, en cuanto a la excepción por "*esfuerzos desproporcionados*", el ICO ha establecido de que solo se aplicara si: (i) el responsable de tratamiento no cuenta con datos de contacto del interesado y no tiene medios razonables para obtenerlos; y (ii) se realice un análisis respecto del esfuerzo requerido para contactar a las personas con el efecto potencial que la extracción de datos tendrá sobre tales personas. Ambos criterios dificultan que se pueda proceder con el cumplimiento de la disposición normativa, en tanto el *Web Scraping* extrae información a grandes volúmenes.

En la práctica, el RGPD del Reino Unido dificulta la extracción de datos personales a través del *Web Scraping*, ya que debe cumplir con uno de los supuestos habilitadores señalados previamente, más aún si se trata de datos especiales. Uno de los casos más famosos en lo que respecta al *Web Scraping* se remonta al año 2010 (Ryanair Ltd v. Billigfluege.de GMBH, 2010), en donde Billigfluege.de, una agencia de viajes alemana extraía información sobre precios y programación de vuelos del sitio web público de Ryanair sin su consentimiento y los publicaba en su plataforma. Si bien la *litis* del problema no fue exactamente por la extracción de datos personales, ello constituyó un antecedente importante al establecer que el *Web Scraping* es ilegal (i) cuando no cumple con los Términos y Condiciones del sitio web y/o (ii) cuando perjudica al tráfico del sitio web. Lo anterior trae como conclusión de que el *Web Scraping* debe ejecutarse con respeto y responsabilidad frente a los demás usuarios y normativa vigente, incluyendo el RGPD del Reino Unido.

Siendo esto así, se evidencia que el Reino Unido se caracteriza por ser uno de los países que ha apostado por las disposiciones del RGPD (2016) y su importancia dentro del ecosistema digital[72]. Además, recientemente el ICO (2024) ha emitido una consulta pública en materia de extracción de datos por *Web Scraping* para fines de entrenamiento de la IA generativa, en donde analiza la base legal de tratamiento por interés legítimo, dado que los otros supuestos habilitadores no calzan en la naturaleza y práctica del *Web Scraping*.

[72] Con fecha 08 de marzo de 2023, se presentó un nuevo Proyecto de Ley (*Data Protection and Digital Informa*tion (No. 2) Bill) que propone una serie de modificaciones relevantes, como la adición de nuevos supuestos para un tratamiento basado en intereses legítimos (ejemplo, se autoriza el tratamiento cuando sea necesario para revelar datos personales a otra persona en respuesta a una solicitud de esta).

Para ejecutar la extracción de datos mediante el *Web Scraping*, el responsable del tratamiento aplicará y superará la prueba de las "*tres partes*" y demostrar que: (i) la finalidad del tratamiento sea legítima (por ejemplo, interés comercial para obtener ganancias); (ii) el tratamiento es necesario para tal fin (por ejemplo, volumen de datos obtenidos mediante *Web Scraping* a gran escala); y (iii) los intereses del individuo no se ven afectados por el interés que se persigue. En cuanto a este último punto, para el ICO (2024) es vital analizar si el titular del dato perderá el dominio de su información, pues la extracción de datos por *Web Scraping* se suele realizar de forma silenciosa, es decir, sin que el titular tome conocimiento de ello.

A fin de ejemplificar lo anterior, se cuenta con el reciente caso de Clearview, en el cual, al igual que la jurisdicción italiana, realizaba tratamiento de datos personales (imágenes) extraídos mediante la técnica del *Web Scraping*. Al respecto, el ICO (2022a) impuso una multa ascendente a £ 7.5 millones de libras esterlinas y una medida de cese de tratamiento de datos dentro del territorio del Reino Unido, junto con la eliminación de dicha información (2022b). Como parte de su análisis, el ICO (2022) señaló que, además de haber incumplido el deber de información, no se había abordado correctamente el punto (iii), es decir, el titular del dato personal desconocía del tratamiento y no existía una expectativa razonable que permitiera realizar tratamiento de datos en tal sentido[73].

De ese modo, el ICO plantea una solución a todos los desarrolladores de *Web Scraping:* que se realice una DPIA. Gracias a esta herramienta, los desarrolladores podrían conocer los riesgos y las posibles soluciones para que no se infrinjan los derechos de los usuarios. A nuestro entender, la ICO va un paso adelante a comparación de sus países vecinos, en tanto ha comprendido el alto impacto de privacidad expuesto cuando se realiza la extracción de datos mediante el *Web Scraping*, así como la alta complejidad y dificultad en la aplicación normativa del RGPD del Reino Unido.

3.1.2 Estados Unidos

[73] En junio del 2022, Clearview AI interpuso un recurso de apelación a la resolución emitida por la ICO ante el Tribunal de primer nivel del Reino Unido argumentando de que no existió una vulneración al RGPD del Reino Unido. En octubre de 2023, el Tribunal resolvió concluyendo de que el procesamiento de datos personales de Clearview IA estaba fuera de la jurisdicción del RGPD (2016) del Reino Unido al brindar servicios a un país exterior - en este caso, Estados Unidos - que no fuese el Reino Unido. Puede revisar la resolución del Tribunal de primer nivel en el siguiente enlace: https://www.bailii.org/uk/cases/UKFTT/GRC/2023/819.html.

La normativa del tratamiento de datos en Estados Unidos presenta diferencias importantes a comparación de la regulación normativa de la Unión Europea. Esto es porque no existe un único cuerpo normativo unificado aplicable a todo el territorio estadounidense, sino normas federales conforme cada Estado y según su tipo, conforme se detalla a continuación:

- Privacy Act (1974), Código 5 USC 552a, (como se cita en Rojas, 2015) sirve para múltiples sectores, como, por ejemplo: informes crediticios (Fair Credit Reporting Act, Public Law 91-508, modificada frecuentemente entre 1996 y 2001), archivos de televisión por cable (Cable Communications Policy Act, 47 USC 521-611, 1994), registros telefónicos (Telephone Consumer Privacy Act 18 USC 2710, 1994);
- Ley de Fraude y Abuso Informático - Código 18 USC 1030 (1986) (Computer Fraud and Abuse Act - CFAA);
- Ley de Privacidad de las Comunicaciones Electrónicas (1986) (*Electronic Communications Privacy Act* - ECPA);
- Ley de la Privacidad de Información Biométrica de Illinois (2008) (*Illinois Biometric Information Privacy Act* - BIPA);
- Ley de Privacidad del Consumidor de California (2018) (*California Consumer Privacy Act* - CCPA*)*;
- Norma de protección de la privacidad en línea de los niños (1998) *(Children's Online Privacy Protection Rule* - COPPA*)*;
- Ley de Derechos de Privacidad de California (CPRA) (2020) (*California Privacy Rights* A*ct* - CPRA).

A pesar de que se presentó un proyecto de ley bipartidista (American Data Privacy and Protection Act, 2022), varios senadores se opusieron al citado proyecto, por lo que no se cuenta con una norma integral a nivel federal. En la práctica, las leyes de cada estado se aplican a la información personal sobre los residentes en dicho territorio, en cuyo caso muchas empresas que operan en Estados Unidos deben cumplir no solo con la legislación federal, sino también con las numerosas normas estatales sobre privacidad.

Se cuenta con un precedente importante en materia de privacidad que fue desarrollado entre 1998 y 2000 a fin de que los responsables de tratamiento de la Unión Europea y Estados Unidos resguarden los datos personales a su cargo, es decir, no se contara con supuestos de fuga de

información o publicación de datos personales sin la debida notificación o permiso. Se trata de los Principios de Puerto Seguro (2000), y que posteriormente fueron revocados por el Tribunal de Justicia Europeo (2015). Dentro de dicho documento se desarrollaron una sucesión de principios significativos sobre el acopio y extracción de datos personales - y aplicables al *Web Scraping* -, tales como:

A) Principio de notificación (*Notice*): establecía que las entidades se obligaban a informar a los particulares sobre los fines con los cuales se recogían y utilizaban sus datos personales, la contactabilidad con ellos para cualquier interrogante o reclamación, los tipos de terceros a quienes se les revelará la información y las alternativas y recursos que la entidad provee a los particulares para restringir su empleo y su propagación.
B) Principio de Seguridad (*Security*): obligaba a que las entidades tomen las precauciones que necesarias y oportunas para eludir pérdida, variación o destrucción de los datos personales extraídos o recopilados.
C) Integridad de los datos (*Data Integrity)*: los datos deben ser imprescindibles para las finalidades de su extracción y/o recopilación.

Si bien dicha relación de Principios se encuentra invalidada para el tratamiento de datos entre la Unión Europea y Estados Unidos[74], el sector privado puede optar por su uso de forma voluntaria, salvo ciertos sectores regulados como la banca. Es importante reconocer el impacto de los Principios de Puerto Seguro en la legislación norteamericana, pues, por ejemplo, Ohio fue el primer estado en aprobar una legislación de puerto seguro de ciberseguridad (Proyecto de Ley 220, 2018).

El contenido de dichos principios advierte que en caso una empresa sufra por una violación de datos de información personal, entonces tiene una defensa positiva si ha "creado, mantenido y cumplido un programa de ciberseguridad por escrito que contenga salvaguardas administrativas, técnicas y físicas para proteger la información personal que se ajuste razonablemente a un marco de ciberseguridad reconocido por la industria" (por ejemplo, las normas PCI-DSS, el Marco NIST, las publicaciones especiales 800-171, 800-53 y 800-53a del NIST, el marco de evaluación de seguridad FedRAMP, HIPAA, GLBA).

[74] A la fecha, Estados Unidos y la Unión Europea vienen negociando la implementación del Marco de Privacidad de Datos UE- EEUU (*EU–US Data Privacy Framework*).

En cuanto a la autoridad competente, no existe una única entidad fiscalizadora del cumplimiento de derechos de los titulares de datos personales en el territorio estadounidense. Sin perjuicio de ello, la Comisión Federal de Comercio (*Federal Trade Comisión* - FTC*)* tiene las facultades para exigir, por ejemplo: (i) la ejecución de medidas razonables de seguridad; (ii) promover contar con políticas de privacidad exactas y transparentes, evitando escenarios de engaño; (iii) resguardar que se cumplan con los principios de autorregulación aplicables del sector; (iv) resguardar que no se viole los derechos de privacidad de los consumidores al extraer, recopilar, utilizar o compartir su información personal. Asimismo, los titulares de datos pueden ejercer sus derechos mediante una acción privada (y acciones colectivas) por violaciones de la privacidad o de seguridad de la información.

Sin perjuicio de la gran cantidad de normativa de privacidad en Estados Unidos, existen dos normas relevantes en la materia: (i) la *CCPA* (2018), que entró en vigencia desde el 2020, inspirada por el RGPD (2016), y (ii) un conjunto de estándares mínimos de privacidad denominado "Marco de Privacidad de APEC" (2005), según lo definido por el Foro de Cooperación Económica de Asia Pacífico (en adelante, "APEC") (Durán, 2015).

Respecto al CCPA (2018), dicha normativa comparándola con el RGPD (2016) define ampliamente a los datos personales, en tanto la protección de la información personal ("*personal information*") se dirige a proteger datos que identifican, se relaciona o son razonablemente posibles de vincularse con un consumidor. Por ello, se incluye toda información confidencial, datos olfativos, el historial de navegación de sitios web, dirección IP, alias y los registros de actividad del usuario.

En relación a la clasificación de los datos sensibles, durante mucho tiempo Estados Unidos no ha establecido su reconocimiento ni regulación. Sin embargo, a partir del CCPA (2018), se comenzó a reconocer dicha categoría que refiere a la seguridad social, permiso de conducir, tarjeta de identificación estatal, pasaporte, datos de acceso a la cuenta, número de cuenta financiera, tarjeta de débito o tarjeta de crédito, creencias filosóficas, ciudadanía, estatus migratorio, información de la niñez, afiliación sindical, y los contenidos de los correos electrónicos y mensajes de texto, salvo que la empresa sea el destinatario de la comunicación. Además, corresponde señalar que las normas de privacidad aplicables a Estados Unidos no incluyen dentro de la categoría de datos sensibles a las opiniones políticas, así como creencias filosóficas. Además, no se incluye el origen racial o étnico como información sensible.

Un tema relevante es que el CCPA (2018) no considera a la información públicamente disponible como dato personal, en tanto entiende que se trata de información puesta a disposición por entidades gubernamentales (Proyecto de Ley No. 375, 2018).

En cuanto a la extracción de datos personales, el CCPA (2018) establece que la empresa podrá realizar la recopilación de un consumidor siempre y cuando sea razonablemente necesario y proporcionado para alcanzar las finalidades objeto de su recopilación o con otra finalidad compatible con el contexto que originó dicha recopilación.

Posteriormente, se emitió el CPRA (2020) que modificó y amplió el contenido de la CCPA (2018). A partir de dicha normativa, distintos estados federales iniciaron a promulgar regulación protegiendo a los datos sensibles, como el estado de Colorado, Connecticut y Utah[75].

Sobre el Marco de Privacidad de APEC (2005), este define a la información personal como "*cualquier información acerca de un individuo identificado o identificable*" (Chan, 2008). Esta normativa aplica a todas aquellas empresas que obtienen más de 25 millones de dólares de ingresos brutos anuales o mantienen información personal de más de 50.000 consumidores, hogares y dispositivos.

Además de dichos cuerpos normativos, se cuenta con el *Video Privacy Protection Act* (1988) que entiende como datos personales a toda aquella información que identifica a una persona –es decir, todos aquellos datos identificables no entran en la definición –; y el *Privacy Act* (PII) (1974), que define como dato personal a aquella información personal identificable.

Por otro lado, actualmente, las Cortes de Estados Unidos han reconocido expresamente que los usuarios de internet tienen interés en la propiedad de su información y datos personales. Ello se ha evidenciado en distintos pronunciamientos, tales como Calhoun v. Google, LLC. (2021), que reconoce el interés de propiedad en la información personal y rechaza el argumento de Google de que la información no es propiedad; y Marriott Int'l Inc. Customer Data Sec. Breach Litig. (2023), en donde se reconoce el valor de la pérdida de propiedad de información personal.

[75] Revisar COLO. REV. STAT. § 6-1-1303(24) (2021); 2023 Conn. Pub. Acts No. 22-15 § 1(27); UTAH CODE ANN. § 13-61-101(32) (West 2023)

Conforme indicamos previamente, la tendencia de Estados Unidos es excluir del concepto de datos personales a la información pública encontrada en los sitios web. Esto se aprecia en el Caso LinkedIn Corp. vs. hiQ Labs Inc. (2019), que se dedicaba a extraer, mediante la técnica del *Web Scraping*, información de perfiles de usuarios públicos alojados en LinkedIn. LinkedIn argumentó su posición en base al CFAA a fin de proteger los intereses de privacidad de sus usuarios de un *scraper* (HiQ Labs, Inc. v. LinkedIn Corp., 2019), dado que hiQ extrajo perfiles públicos del sitio web de LinkedIn para producir el análisis llamado "Keeper", que permitía identificar a potenciales candidatos de ser reclutados[76]. Por ello, LinkedIn argumentó que la extracción de datos de hiQ y el análisis "Keeper" ponía en peligro la privacidad de sus usuarios que utilizaban LinkedIn dado que, si bien solo se *scrapeaba* perfiles públicos, muchos de los usuarios no desean que sus empleadores conozcan que están buscando nuevo empleo[77].

En este caso, el Tribunal de Distrito de los Estados Unidos para el Distrito Norte de California decidió que la información alojada en LinkedIn no necesitaba de algún tipo de consentimiento al ser una página web pública, considerando además que los usuarios asumen el riesgo de que un tercero pueda visualizar su perfil.

Otro caso relevante es el de Clearview AI Inc. (2020), en donde dicha empresa extraía datos personales y sensibles (p.e.,fotografías) de sitios webs y redes sociales, como LinkedIn y Facebook, con el propósito de ser utilizados por entidades gubernamentales. Clearview AI Inc. (2020) alegaba que los individuos no tenían derecho a la privacidad, pues estos mismos publican en internet dicha información personal. Asimismo, en el caso de Cox Corp. v. Cohn (1975), donde una víctima de violación demandó a una estación de televisión por transmitir su nombre. La Corte concluyó que los intereses en la privacidad se desvanecen cuando la información involucrada ya aparece en el registro público (Cox Corp. v. Cohn, 1975, 494). El mismo razonamiento se aplicó en el caso de Florida Star v. B.J.F. (1989) cuando un periódico publicó el nombre de una víctima de violación después de obtenerlo de un informe policial disponible públicamente. Aunado a ello, en el caso Feits Publications, Inc. Vs. Rural Telephone Service Co. (1991), la Corte Suprema de Estados Unidos sostuvo que extraer y republicar información pública, como las listas del teléfono, se encuentran permitidos.

[76] Ver Id. at 991.
[77] Ver Id. at 994

Entonces, podemos concluir que en Estados Unidos se sigue la regla de "*no consentimiento previo y expreso sobre información pública, además de los demás derechos de protección de datos personales*" al determinar que los individuos carecen de expectativas razonables de privacidad en espacios públicos (Bedi, 2017).

Siendo que el razonamiento del legislador estadounidense cuenta con una postura marcada y establecida, la doctrina advierte que existen intereses de protección de la privacidad muy fuertes en la información personal pública y, por ello, han articulado varias teorías al respecto:

1. Los daños a la privacidad de la información personal pública: sostiene que los daños a la privacidad son independientes de si la información es inicialmente pública o privada, pues la sola extracción de los datos genera daños emocionales. El profesor J. Solove (2006) sostiene que la extracción de información ilegal "crea sentimientos de ansiedad e incomodidad". Asimismo, el procesamiento de dicha información puede generar daños. Por ejemplo, los investigadores analizaron tuits públicos para identificar a los usuarios con problemas de salud mental (Reynolds, 2020). El *Web Scraping* también genera daños a la seguridad, pues actualmente, los raspadores han sufrido violaciones de datos (Scroxton, 2020). Además, la falta de proporcionar un aviso de los usuarios (el profesor Solove lo denomina como "exclusión") de que se encuentran expuestos a ser víctimas del *Web Scraping* puede ser en sí misma un daño (Solove, 2006).
2. Privacidad debido a la oscuridad: La noción de que cuando es poco probable que nuestras actividades o información se encuentren, vean o recuerden, es, hasta cierto punto, segura (Hartzog, 2019). Además, podemos esperar razonablemente conservar los intereses de privacidad en tal información incluso si es técnicamente pública, en el sentido que no se espera que los demás revisen en la información del otro (Hartzog, 2019). Asimismo, el tiempo es un elemento oscurecedor, pues hay pocas probabilidades de acceder a, por ejemplo, redes sociales antiguas como *Myspace* y obtener a partir de ahí datos personales.
3. Privacidad debido a la confianza en sitios web y otros usuarios: Según Balkin (2016), la relación sitio web/usuario es una de confianza, al igual que la relación abogado-cliente. En consecuencia, los sitios web son "fiduciarios de la información" y tienen "deberes especiales con respecto a información personal que obtienen dentro de la relación" con sus usuarios (Khan y Pozen, 2019). Esto incluye protegerlos contra el

Web Scraping no autorizado, incluyendo la información personal pública (disponible en su *website*).

4. Publicar públicamente no es un consentimiento implícito: Consiste en que la normativa de privacidad, como el RGPD (2016), reconoce que se requiere necesariamente de un "acto afirmativo claro" como el consentimiento, por lo que un consentimiento implícito, como "casillas previamente marcadas o inactividad", resulta insuficiente.

En suma, existen fuertes intereses de protección de la privacidad en la información personal pública. Es así que el *Web Scraping* de terceros no autorizado pone en riesgo significativo la privacidad de los usuarios. Frente a este contexto y a pesar de la ponderación entre la privacidad y la información pública, podemos señalar que el panorama regulatorio actual es el siguiente:

1. El sitio web puede interponer un reclamo ante la CFAA (1986)

En materia de consumidor, el sitio web puede argumentar que ha colocado en sus términos de servicio la prohibición de la técnica del *Web Scraping* (Riley, 2018). De ese modo, la discusión gira en torno a si los Términos y Condiciones requieren (a) "real o conocimiento constructivo" (Nguyen v. Barnes & Noble Inc, 2014) y (b) si son fácilmente visibles dentro del sitio web.

Por un lado, sobre el conocimiento real o constructivo, en un caso de extracción de datos, se cuenta con precedentes que indican que basta con que se incluya una disposición dentro de los términos y condiciones que prohíba el *Web Scraping* (DHI Grp. Inc. v. Kent, 2017). Asimismo, se toma en cuenta la ubicación de la prohibición de no extraer para el conocimiento constructivo.

Por otro lado, la visibilidad, los términos y condiciones contenidos en hipervínculos en la parte inferior de una página web generalmente no son suficientemente visibles para los usuarios. Ejemplos claros de cómo se pueden redactar los términos y condiciones del sitio web son los términos de uso de Zillow que prohíben consultas automatizadas, específicamente el *Web Scraping*[78]. Por otro lado, los términos de uso de Etsy establecen de manera expresa que los

[78] Terms of Use, ZILLOW, https://www.zillow.com/corp/Terms.htm [https://perma.cc /7AJP-FFU9] (última visita: 19 de mayo de 2024). Zillow, however, offers direct downloads of certain research data.

usuarios se comprometen a no ”scrapear”, ”raspar o ”arañar” ninguna página de su sitio web[79]. Existen diversas páginas web como: Twitter[80], Facebook[81] y LinkedIn[82], entre otras que prohíben el *Web Scraping* y los bots en sus términos y condiciones.

Esta postura sigue la tendencia doctrinaria de responsabilizar a los titulares de los sitios web, pues, además de colocar la prohibición en sus términos y condiciones, son capaces de monitorear el *Web Scraping* mediante la inspección del tráfico web (Xiao,2021).

2. Los usuarios y las agencias reguladoras pueden hacer valer los reclamos de la ley estatal de privacidad de datos (por ejemplo, CCPA (2018) y BIPA (2008))

Los Estados de California y Vermont cuentan con regulación sobre los corredores de datos, entendidos como empresas que extraen venden a terceros información personal de un consumidor con quien no cuenta una relación directa (Código Civil de California, 1798.99.80(d)). Tal normativa permite corregir la asimetría de información en la relación *scraper* - usuario, pues los usuarios desconocen que intermediarios cuentan con sus datos personales, en cuyo caso amerita que puedan ejercer su derecho de cancelación.

3. Las agencias reguladoras pueden hacer cumplir la normativa de protección de datos estatales

Las dos normas principales de privacidad a nivel Estado son: la CCPA (2018) y BIPA (2008) de Illinois. Ambos cuerpos normativos otorgan poder al usuario en lugar del titular del sitio web (sobre cómo se utiliza su información) y le otorga derechos como derecho a recibir notificación de la extracción de sus datos frente a la “información personal” extraída por entidades reguladas (Código Civil de California, 1798.100(b)) que también se aplican a información personal pública.

[79] Términos de uso – Las normas de la casa, ETSY, https://www.etsy.com/legal/terms-ofuse/ [https://perma.cc/P58A-EX5B]. Aunque Etsy ofrece una API, sus términos también contienen una disposición que prohíbe el scraping automatizado y los bots. Véase Condiciones de uso de ETSY, https://www.etsy.com/legal/api [https://perma.cc /58AG-98CQ].
[80] X Terms of Service, X, https://x.com/en/tos (Última visita: 18 de mayo de 2024) (“crawling or scraping the Services in any form, for any purpose without our prior written consent is expressly prohibited”).
[81] Condiciones del Servicio, Facebook (última visita el 18 de mayo de 2024), https://www.facebook.com/terms.php. ”No puedes acceder a datos desde nuestros Productos ni recopilarlos usando medios automatizados (sin nuestro permiso previo), ni intentar acceder a datos si no tienes permiso para hacerlo."
[82] User Agreement, LinkedIn https://www.linkedin.com/legal/user-agreement (última visita el 18 de mayo de 2024) (“You agree that you will not: . . . [d]evelop, support or use software, devices, scripts, robots, or any other means or processes (including crawlers, browser plugins and add-ons, or any other technology or manual work) to scrape the Services or otherwise copy profiles and other data from the Services”).

En cuanto al deber de informar, la falta de notificación a los usuarios en la extracción de datos en la legislación de Estados Unidos es la regla general. Como excepción, solo los residentes de Illinois conforme a la BIPA (2008), se les exige que las empresas obtengan el consentimiento de los titulares de los datos personales antes de extraer sus datos biométricos[83].

En esa línea, la CCPA (2018) requiere que las empresas notifiquen a los consumidores sobre las categorías de información que planean recopilar y sus finalidades, siempre y cuando se vaya a comercializar dicha información. Es decir, les brinda a los consumidores los siguientes derechos: (i) el derecho a conocer la información personal que una empresa recopila sobre ella y cómo se utiliza y comparte; (ii) el derecho a eliminar la información personal recopilada de ellos (con algunas excepciones como la información publicada en páginas públicas); (iii) el derecho a optar por no vender o compartir su información personal ("*do-not-sell-or-share my information*") mediante un enlace dentro del sitio web; y (iv) que se cuente con un método en línea y un número de teléfono gratuito para ejercer sus derechos. Desde el 01 de enero de 2023, con la Proposición 24 (2023), la CPRA (2020) ha incluido nuevos derechos al CCPA (2018) como: (v) derecho a corregir información personal inexacta que una empresa tenga sobre ella; y (vi) el derecho a limitar el uso y divulgación de información personal sensible extraída sobre ellos.

Es importante recalcar que, como mencionamos previamente, la CCPA (2018) no se aplica a la información que esté a cargo de las entidades gubernamentales. Esto no ocurre con la BIPA (2008), que no hace ninguna excepción para la información disponible públicamente.

De otro lado, la COPPA (1998) exige el consentimiento de los padres antes de extraer cualquier información personal de niños menores de 13 años. Asimismo, la CCPA (2018) exige que una empresa obtenga el consentimiento explícito antes de la venta de cualquier información personal sobre un consumidor que la empresa tenga "*conocimiento real*" de que es menor de 16 años, y cuando el consumidor sea menor de 13 años, se requiere la autorización expresa paterna.

[83] Comparar 740 ILL. COMP. STAT. 14/15(b)(1)-(2) (2008) (que exige un aviso previo a la recogida) con el Código de California, CAL. CODE REGS. tit. 11, § 999.305(d) (2020) (exención de los rascadores de la obligación de notificación previa a la recogida).

En consecuencia, mientras que la CCPA (2018) exime a los *scrapers* de dar aviso al titular de los datos (salvo se comercialice dichos datos personales), la BIPA (2008) requiere de la notificación por cada vez que se extraen los datos personales. Para los *scrapers*, esto significa: (a) derechos de acceso: deberán encontrarse preparados para responder a las solicitudes de los consumidores con respecto a su información personal extraída a través de actividades de *Web Scraping*, (b) derecho de eliminación: garantizar que existan mecanismos para que los consumidores ejerzan este derecho y (c) derecho de exclusión voluntaria: los consumidores tienen derecho a optar por no participar como usuarios en cualquier venta de su información personal.

Asimismo, la CCPA (2018) no incluye la conducta de extraer información de un sitio web en su definición de recopilación; sin embargo, una implementación de la CCPA (2018) o el Reglamento de la CCPA (2018) especifica en la sección 999.305 (d) que en caso una empresa no recopile datos personales directamente del consumidor, como es el caso del *Web Scrapimg*, entonces no requiere proporcionar un aviso de recolección si no comercializa tales datos personales. En su exposición de motivos, el fiscal general de California explicó que las empresas que extraen información personal indirectamente de los consumidores no es factible proporcionar un aviso "en o antes del punto de recolección" (California Office of the Attorney General, s.f.); por ello, se exime a los *scrapers* del requisito de notificación, salvo comercialicen la información.

En conjunto de lo revisado previamente, podemos advertir que, dentro de la jurisdicción de Estados Unidos, no existen leyes federales que prohíban el *Web Scraping* como tal siempre que (i) los datos estén disponibles públicamente y, además, (ii) que no se ocasione algún tipo de daño al sitio web. Sin perjuicio de ello, consideramos que la notificación es fundamental en el contexto del *Web Scraping*, en tanto existe una relación indirecta entre usuarios y *scrapers*, en donde los usuarios no conocen la existencia de estos últimos[84] y actúan secretamente. No obstante, si los usuarios supieran que los *scrapers* extraen información pública, estos podrían optar por hacer que su información sea privada (Susser, 2019).

Actualmente, la discusión en torno al *Web Scraping* ha llegado a un siguiente nivel: las compañías de entrenamiento de inteligencia artificial utilizan la técnica del *Web Scraping* para

[84] Usuarios de Twitter desconocen que sus tuits públicos pueden ser utilizados por investigadores, por ejemplo.

la práctica de sus sistemas, pudiendo perjudicar en gran escala los derechos de protección de datos personales. En este caso, resulta discutible si, al amparo del CCPA (2018), la extracción de datos de capacitación para el aprendizaje automático requiere de notificación (Garhart, 2020). En contraria, la BIPA (2008) no distingue entre recolección directa e indirecta, pues todos los raspadores deben avisar antes de la extracción de datos[85].

Un interesante caso sobre extracción de datos para entrenamiento de IA es lo que ocurre en el caso *P.M et al. V. OpenAI LP* (2023). El pasado 28 de junio de 2023, dentro del Tribunal de Distrito de los Estados Unidos para el Distrito Norte de California, se presentó una demanda contra OpenAI por recopilar y extraer datos personales[86] a través del *Web Scraping* de forma ilegal; ello con el propósito de alimentar y desarrollar los productos pertenecientes a dicha compañía para su futuro uso y comercialización.

Además de exigir transparencia y responsabilidad en el uso de los datos personales extraídos, se le solicitaba a OpenAI que permita a los usuarios de internet a tener el control sobre sus datos personales, optando por no participar en su extracción de manera indiscriminada mediante el *Web Scraping*. A pesar de no contar con el debido consentimiento del titular, OpenAI alimentaba sus modelos de IA con todos los datos derivados de las interacciones de los usuarios -cada clic, entrada, pregunta, uso, cada movimiento, pulsación de tecla, búsqueda, geolocalización del usuario como datos de entrenamiento. Se estima que OpenAI recopiló alrededor de 300.000 millones de palabras de diversas fuentes, como libros, artículos, sitios web y mensajes. A la fecha, las partes se encuentran intercambiando información y pruebas dentro del proceso, y podrían ir a juicio si es que no llegaran a un acuerdo.

Es así que podemos concluir que en Estados Unidos no se prohíbe el *Web Scraping* con la información pública disponible mientras no se haga efectivo un daño contra la misma página

[85] La BIPA exige a todos los recopiladores de datos biométricos que notifiquen y reciban el consentimiento antes de recopilar información biométrica. 740 ILL. COMP. STAT. 14/15(b) (2008) ("No private entity may collect, capture, purchase, receive through trade, or otherwise obtain a person's or a customer's biometric identifier or biometric information, unless [notice is given and consent is received].").

[86] Los datos personales sometidos a recopilación y extracción serían (1) todos los datos introducidos en los productos de OpenAI; (2) información de la cuenta que los usuarios introducen al registrarse; (3) nombre; (4) datos de contacto; (5) credenciales de inicio de sesión; (6) correos electrónicos; (7) información de pago para usuarios de pago; (8) registros de transacciones; (9) datos de identificación extraídos de los dispositivos y navegadores de los usuarios, como direcciones IP y ubicación, incluida la geolocalización de los usuarios; (10) información de redes sociales; (11) datos de registro de chat; (12) datos de uso; (13) análisis; (14) cookies;15 (15) pulsaciones de teclas; y (16) búsquedas escritas, así como otros datos de actividad en línea. incluidos, entre otros, ubicaciones de usuarios y datos relacionados con imágenes obtenidos a través de Snapchat, información financiera de usuarios a través de Stripe, gustos y preferencias musicales a través de Spotify, patrones de usuarios y análisis de conversaciones privadas a través de Slack y Microsoft Teams e incluso información sanitaria privada obtenida a través de la gestión de portales de pacientes como MyChart.

web. No obstante, se promueve una cultura de autorregulación, en la que los titulares de los sitios web reivindiquen los intereses de privacidad de los usuarios al contar con mejores y mayores herramientas técnicas contra el *Web Scraping* ilegal. Tal iniciativa se ejecutaría dentro del marco de las leyes de privacidad de datos como CCPA (2018), por ejemplo, al permitir otorgar poder al usuario sobre el control de la extracción de su información personal, incluyéndose la opción de otorgar un consentimiento expreso y previo por parte del titular de datos personales. A la fecha, se cuenta con grandes comunidades profesionales dedicadas a impartir el adecuado uso ético y legal del *Web Scraping*[87], promoviendo la cultura de autorregulación y ética plena al momento de extraer información.

3.1.3 Latinoamérica

Tal como indica Rojas (2014), en América Latina no existe un tratado regional que regule el derecho a la protección de datos personales de forma uniforme y conjunta, como es el caso de la Unión Europea. Sin embargo, existen iniciativas regulatorias suscritas por distintos países de la región con el propósito de desarrollar normativa alineada y necesaria para garantizar una adecuada protección de datos personales, como es el caso de la Red Iberoamericana de Protección de Datos, originada a raíz del Encuentro Iberoamericano de Protección de Datos (EIPD) en el 2003.

Tal Red se encuentra compuesta por doce países, en los que se incluyen países de Latinoamérica, como México, Perú, Argentina y Colombia. Uno de los puntos claves fue la emisión de los "Estándares de Protección de Datos Personales para los Estados Iberoamericanos" (2017), que cuentan con el respaldo de la Comisión Europea y representan un modelo de referencia sobre la regulación futura de datos personales.

Dentro de su contenido se establece un artículo de alta inspiración del RGPD (2016) – y trascendencia para la legitimidad de tratamiento mediante el *Web Scraping* –, pues se permite el tratamiento si es necesario para la satisfacción de intereses legítimos perseguidos por el responsable en caso no perjudiquen los intereses, derechos o libertades fundamentales del titular del dato, siendo un caso especial cuando el titular sea niño, niña o adolescente[88].

[87] Por ejemplo, existe el Web Data Extract Summit 2024 que se realizará en octubre del 2024. Para mayor información, podrá acceder al siguiente enlace: https://www.extractsummit.io/.

[88] 11. Principio de legitimación
11.1. Por regla general, el responsable solo podrá tratar datos personales cuando se presente alguno de los siguientes supuestos:

Asimismo, se cuenta con el Marco de Privacidad APEC (2005) suscrita por aquellos países que forman parte de la APEC, tales como Perú y México. El documento resalta la importancia del flujo de información entre las economías de mercado y las medidas de seguridad de la información que deben tomarse en cuenta a fin de evitar contingencias. Se advierte el principio de "Prevención de daños"[89] en el que se reconoce los intereses del individuo a las expectativas legítimas de privacidad, siendo que todas las medidas deben ser proporcionales a la gravedad del daño que amenaza la recopilación y extracción de datos personales. Consecuentemente, el Marco de Privacidad (2005) contiene el principio de "Limitación de recopilación"[90] en el que advierte que toda recopilación deberá ser limitada por los fines de que lo motivan y deberá ser obtenida por medios lícitos con previo consentimiento o notificación de la persona afectada.

Sin perjuicio de los acuerdos y marcos regionales que han pretendido establecer medidas estándar para los países de la región, es de conocimiento que no se cuenta con una vasta regulación sobre la materia. No obstante, la (no tanta) existente sigue el patrón e influencia del RGPD (2016) para cada cuerpo normativo y pronunciamiento. Por ello, se ha identificado a los principales países de la región en materia de protección de datos con el propósito de dilucidar si es que se han tomado medidas o posturas en cuanto a la extracción de datos mediante el *Web Scraping*.

A) Argentina

Argentina resulta un país particular de la región al ser pionera en materia de protección de datos personales. En el artículo 43 de la Constitución Federal, se señala que, por ejemplo, una persona podrá interponer acción para conocer la finalidad por la que se conservan los datos personales (Constitución de la Nación Argentina, 1994). Si bien dicho reconocimiento no crea un derecho

i. El tratamiento sea necesario para la satisfacción de intereses legítimos perseguidos por el responsable o por un tercero, siempre que sobre dichos intereses no prevalezcan los intereses o los derechos y libertades fundamentales del titular que requiera la protección de datos personales, en particular cuando el titular sea niño, niña o adolescente. Lo anterior, no resultará aplicable a los tratamientos de datos personales realizados por las autoridades públicas en el ejercicio de sus funciones.

[89] Principles

I. Preventing harm

14. Recognizing the interests of the individual to legitimate expectations of privacy, personal information protection should be designed to prevent the misuse of such information. Further, acknowledging the risk that harm may result from such misuse of personal information, specific obligations should take account of such risk, and remedial measures should be proportionate to the likelihood and severity of the harm threatened by the collection, use and transfer of personal information.

[90] III. Collection Limitation

18. The collection of personal information should be limited to information that is relevant to the purposes of collection and any such information should be obtained by lawful and fair means, and where appropriate, with notice to, or consent of, the individual concerned

constitucional a la protección de datos, resulta un avance como fundamento para su (posible) futuro reconocimiento.

Asimismo, en cuanto a la doctrina autorizada, Peyrano (2007) refiere que la regulación argentina en materia de privacidad se rige bajo la estructura y el contenido del RGPD (2016), siendo además influenciada, específicamente, por la normativa española. De hecho, la Comisión Europea la ha considerado como una regulación adecuada al seguir las normas internacionales. Actualmente, se cuenta con la Ley de Protección de Datos Personales, Ley No. 25.326 (en adelante, "Ley Argentina") (2000), siendo la primera generación de legislaciones en América Latina sobre la materia.

La Ley Argentina (Ley No. 25.326, art. 2) es una ley federal que busca proteger a los datos públicos o privados en el territorio, incluyendo a las transferencias internacionales para su tratamiento. Dicha norma define a los datos personales como aquella información de cualquier tipo referido a personas físicas o de existencia ideal, sean determinadas o determinables[91]. Es decir, abarca también a la protección de personas jurídicas. Para complementar, el objeto de la norma incluye a los datos personales que se encuentran dentro de registros, bancos de datos, *archivos* u otros medios técnicos de tratamiento de datos que otorguen informes (Ley No. 25.326, art.1)[92].

Asimismo, define a los "datos sensibles" como los datos personales vinculados al origen racial y étnico, opiniones políticas, creencias religiosas, filosóficas o morales, sindicalización e información referente a la salud o sexualidad. En cuanto al tratamiento de datos, se requiere del consentimiento del interesado, que deberá ser libre, expreso e informado y constar por escrito o en otra forma equivalente, salvo que, se hayan obtenido de fuentes de acceso público no restringido.

[91] Artículo 2° - Definiciones
A los fines de la presente ley se entiende por:
- Datos personales: información de cualquier tipo referida a personas físicas o de existencia ideal determinadas o determinables. (...)

[92] Artículo 1° - (Objeto). La presente ley tiene por objeto la protección integral de los datos personales asentados en archivos, registros, bancos de datos, u otros medios técnicos de tratamiento de datos, sean éstos públicos, o privados destinados a dar informes, para garantizar el derecho al honor y a la intimidad de las personas, así como también el acceso a la información que sobre las mismas se registre, de conformidad a lo establecido en el artículo 43, párrafo tercero de la Constitución Nacional. Las disposiciones de la presente ley también serán aplicables, en cuanto resulte pertinente, a los datos relativos a personas de existencia ideal. En ningún caso se podrán afectar la base de datos ni las fuentes de información periodísticas.

Por otro lado, el 03 de diciembre de 2001, se publicó el Decreto No. 1558 que reglamenta la Ley Argentina (en adelante, "Decreto 1558"), que reglamentó la acción constitucional de hábeas data (incorporada en la Constitución Nacional con la reforma de 1994).

Otro cuerpo normativo relevante es el Código Civil y Comercial Ley No. 26994, vigente desde el 1º de agosto de 2015, siendo que el artículo 53 regula el derecho a la imagen y a la voz de una persona. El citado artículo dispone la necesidad de consentimiento de las personas para captar o reproducir la imagen o su voz, salvo cuando se trate de interés público, en cuyo caso se deben tomar medidas suficientes para evitar un daño innecesario (Congreso de la Nación Argentina, 2014).

Como autoridad sobre la materia, existe la Agencia de Acceso a la Información Pública (AAIP) en calidad de una autoridad autónoma. Al respecto, se ha emitido pronunciamientos relevantes concernientes al acceso de las fuentes públicas, en tanto, mediante Dictamen de la DNPDP No. 32/2009, advirtió que no toda la información de carácter personal asentada en un registro público puede difundirse o cederse por pertenecer a una base de datos pública (2009). En esa misma línea se pronunció mediante el Dictamen de la DNPDP No. 009/04 (2004).

Por otro lado, el Centro de Ciberseguridad de la Jefatura de Gabinete de Argentina se ha pronunciado positivamente sobre el *Web Scraping,* identificándolo como una herramienta útil para la extracción de información. No obstante, es imprescindible que los usuarios verifiquen que los sitios web sean transparentes y cumplan con la normativa de protección de datos personales, además de corroborar si se prohíbe o no el *Web Scraping* (Centro de Ciberseguridad, 2024).

En cuanto a jurisprudencia relevante, no existen casos judiciales o administrativos de *Web Scraping*. Sin embargo, es común que esta práctica sea utilizada dentro del país, por lo que Argentina ha suscrito la Declaración conjunta sobre extracción de datos y protección de datos (Information Commissioner's Office, 2023*)*, en la cual se reconoce al *Web Scraping* como un riesgo potencial para la privacidad de los usuarios y destaca la responsabilidad de la protección de datos de los titulares a los operadores de sitios web y redes sociales a través de medidas técnicas de prevención frente a tal técnica.

B) México

México constituye un país particular al incluir el derecho de la protección de datos personales de forma expresa en su Constitución[93], a través de una reforma constitucional en el 2009. El reconocimiento incide en los derechos de los titulares frente al tratamiento de sus datos, es decir, los derechos de acceso, rectificación, cancelación y oposición (en adelante," Derechos ARCO") basándose en un marco de principios que debe existir en todo tratamiento de datos. Es así que, con tal reconocimiento, se otorga una garantía individual llegando a constituirse un derecho fundamental y autónomo.

Luego de dicho reconocimiento constitucional, en el 2010 se publicó la Ley Federal de Protección de Datos Personales (en adelante, "LFPDPPP"), la primera ley federal sobre la materia, para posteriormente emitirse su Reglamento (2011). La LFPDPPP tiene, como parte de sus finalidades, el establecer las obligaciones para los responsables del tratamiento a fin de que únicamente realicen el tratamiento en función a los fines que han motivado la recopilación o extracción de dichos datos personales.

La regulación mexicana entiende el concepto de dato personal como aquel que se encuentra unido a una persona física y, por ello, merece protección al titular del dato, no el dato en sí. En base a ello, la LFPDPPP (2010) define el concepto de "datos personales" como cualquier información concerniente a una persona física identificada o identificable. Asimismo, existe una categoría especial denominada datos personales sensibles, que, como ya se definió anteriormente en las distintas regulaciones vigentes, son aquellos datos que afectan la esfera más íntima de su titular o cuya utilización indebida pueda causar discriminación o conlleve un riesgo grave para esto. Por ejemplo: datos de origen racial o étnico, estado de salud – presente y futuro-, entre otros.

La aplicación de la LFPDPPP (2010) se entiende a todos aquellos tratamientos que (i) se realicen en territorio mexicano, (ii) sean realizados por un encargado del tratamiento, independientemente de su ubicación, si el tratamiento se realiza por cuenta de un responsable del tratamiento establecido en México, (iii) sean realizados por o en nombre de un responsable

[93] Artículo 16
(...)
Toda persona tiene derecho a la protección de sus datos personales, al acceso, rectificación y cancelación de los mismos, así como a manifestar su oposición, en los términos que fije la ley, la cual establecerá los supuestos de excepción a los principios que rijan el tratamiento de datos, por razones de seguridad nacional, disposiciones de orden público, seguridad y salud públicas o para proteger los derechos de terceros. (...)

de datos no establecido en México, cuando la legislación mexicana sea aplicable en virtud de la ejecución de un acuerdo o de la adhesión de México a un convenio internacional o (iv) sean realizados en territorio mexicano, por cuenta de un responsable de datos no establecido en territorio mexicano, salvo que dicho tratamiento sea únicamente para fines de tránsito. Del mismo modo, de acuerdo con la LFPDPPP únicamente aplica a personas y entidades privadas, mas no a las entidades del sector público, así como también a datos destinados a una empresa (personas naturales que actúen en nombre de la persona jurídica en el marco de representación) o que actúen como profesionales o comerciantes (2010).

Por otro lado, el Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (en adelante, "INAI") es el ente rector encargado del cumplimiento de la LFPDPPP y su Reglamento (2010), además de emitir lineamientos y pronunciamientos vinculantes en materia de protección de datos personales. Al respecto, el INAI se caracteriza por estar a la vanguardia de los nuevos fenómenos tecnológicos existentes en su territorio, por lo que ha emitido una serie de guías orientadoras, tales como las "*Recomendaciones para mantener segura la privacidad y datos personales en el entorno digital*" (INAI, 2018).

Además de ello, la LFPDPPP (2010) obliga a que el responsable de tratamiento designe a un Delegado de Protección de Datos para que pueda atender las solicitudes de los Derechos ARCO, así como supervisar el cumplimiento de la LFPDPPP (2010) y su Reglamento (2010) dentro de la organización.

Al igual que otras jurisdicciones, los datos personales deben extraerse con fines determinados, explícitos y legítimos, siendo que no debe existir un tratamiento incompatible con dichos fines, así como también ceñirse a los principios de consentimiento, información, responsabilidad y legalidad. Sobre el particular, los interesados tienen derecho a una expectativa razonable de privacidad en el tratamiento de sus datos personales, incluyendo que se debe disponer de un Aviso de Privacidad antes de su tratamiento.

El contenido del Aviso de Privacidad variará en función si es que los datos personales han sido obtenidos de forma directa o indirecta del interesado, así como el contexto y espacio:

- Aviso de Privacidad exhaustivo: obligatorio cuando los datos personales se obtienen personalmente del interesado, por ejemplo, en una entrevista física. En tal Aviso de

Privacidad deberá cumplirse de forma completa y expresar el deber de información, es decir, detallar los aspectos referidos a informar cómo se tratarán los datos personales.

- Aviso de Privacidad simplificado: cuando los datos se obtienen directamente del interesado, por ejemplo, al registrarse en una cuenta de un sitio web o una llamada. En tal tipo de Aviso de Privacidad se incluirá la identidad y dirección del responsable de tratamiento, las finalidades (primarias y secundarias) de tratamiento, la vía para ejercer sus derechos ARCO y cómo poder acceder al Aviso de Privacidad exhaustivo.
- Aviso de Privacidad abreviado: cuando el espacio para detallar los aspectos del deber de información es limitado y los datos personales recogidos son mínimos, por ejemplo, en un SMS o un pequeño formulario. En tal tipo de Aviso de Privacidad se incluirá la identidad y dirección del responsable de tratamiento, las finalidades de tratamiento (sin distinguir primarias y secundarias) y la vía para ejercer sus derechos ARCO.

Tal Aviso de Privacidad, puede entregarse en formato impreso, digital, visual o sonoro en cualquier otra tecnología, siendo además que debe ser claro y con un lenguaje comprensible.

En cuanto al consentimiento, la regla en general es que el consentimiento sea implícito o tácito, salvo que se traten de datos sensibles (en el que se requiere que sea expreso y por escrito), datos financieros o patrimoniales, o cuando la normativa lo exija específicamente. (LFPDPPP, 2010). El consentimiento implícito se obtendrá cuando el interesado ha sido informado del Aviso de Privacidad y no se ha opuesto a su tratamiento.

Respecto al tratamiento de datos personales mediante *Web Scraping*, a la fecha no se cuenta con pronunciamientos judiciales o administrativos que nos permita dilucidar o exponer el razonamiento del legislador mexicano sobre tal técnica; no obstante, México suscribió la Declaración conjunta sobre extracción de datos y protección de datos (Information Commissioner's Office, 2023) junto con otros países, en cuyo caso entenderíamos que el INAI sí ha identificado los riesgos del tratamiento de datos mediante el *Web Scraping* y, conforme se detalla en la citada Declaración, los pasos a cumplir para no vulnerar la normativa de protección de datos mexicana.

C) Colombia

En Colombia, el derecho de protección de datos tiene su fundamento en el Artículo 15 de la Constitución Política, en la que se reconoció por primera vez el derecho de habeas data (Constitución Política de Colombia, 1991). Así, el Habeas Data fue desarrollado en la Ley Estatutaria No. 1266 de 2008 y sus decretos reglamentarios (Decreto No. 1727 de 2009 y Decreto No. 2952 de 2010), orientados únicamente a la protección de los datos comerciales y financieros, por lo que no garantizaba de manera integral y expresa el derecho a la protección de datos personales.

De esta manera, mediante la Ley 1266 de 2008 (2008), se expidió el Régimen General de Protección de Datos Personales, en donde se establecieron ciertos principios rectores de la protección de datos, además de los derechos del titular de los datos y las obligaciones del responsable del tratamiento. Dicha norma daba énfasis al tratamiento de los datos bancarios, crediticios, comerciales o financieros.

Posteriormente, se emitió la Ley No. 1581 de 2012 (2012) ("Ley No. 1581") que regula todo el tratamiento de datos personales en el territorio colombiano registrados en cualquier base de datos (sea público o privado). En la citada norma se regulan aspectos como la autorización del titular de la información para tratar sus datos personales, las políticas de tratamiento aplicables a los responsables y encargados de tratamiento, la clasificación de datos sensibles, entre otros. Luego, se promulgó el Decreto Reglamentario parcial No. 1377 que reglamentó la Ley No. 1581 (2013) ("Decreto No. 1377").

La Ley No. 1581 (2012, artículo 3 c)) define al dato personal como aquella información vinculada o que pueda derivarse a una o más personas naturales determinadas o determinables. En este caso, se puede equiparar el término "determinable" por "identificable". Lo interesante es que no tiene que ser información relacionada a aspectos de la vida privada o intimidad, pues dicha información no se circunscribe a ningún aspecto particular de la persona (Remolina, 2015). Las regulaciones sobre tratamiento de datos personales no son iguales que las referidas al derecho a la intimidad ni se limitan únicamente a aquella información referida solo a la vida privada de las personas.

Asimismo, existe un precedente interesante de la Corte Constitucional de Colombia (C-748 de 2011, 2011), en el que se extendió el tratamiento de datos personales a las personas jurídicas,

en tanto, bajo razonamiento del legislador, es legítima la referencia a la protección de las personas jurídicas cuando se vulneren derechos de las personas que la conforman.

Cabe resaltar que en Colombia los datos personales pueden ser clasificados (Ley 1266, 2008) en:

- Datos públicos: calificados como tales según los mandatos de la ley o de la Constitución Política (Ley 1266, 2008, artículo 3, literal f).
- Entre los datos públicos se encuentran los datos contenidos en documentos públicos, sentencias judiciales ejecutoriadas que no estén sometidos a reserva y los relativos al estado civil de las personas (Ley 1266, 2008, artículo 3, literal f).
- Datos semiprivados: no tienen naturaleza íntima, reservada ni pública y cuyo conocimiento o divulgación puede interesar no solo a su titular, sino a cierto sector o grupo de personas o a la sociedad en general, como el dato financiero y crediticio de actividad comercial o de servicios (Ley 1266, 2008, artículo 3, literal g).
- Datos privados: son los datos que, por su naturaleza íntima o reservada, solo es relevante para el titular (Ley 1266, 2008, artículo 3, literal g). Los datos privados solo afectan al titular, por lo que dicha información no debe ser observarse o tener injerencias indebidas por ningún órgano público o privado.

En cuanto a lo que se entiende por **dato público**, el Decreto No. 1377 (2013, art. 3. 2)) lo define como aquel dato que no sea privado, semiprivado o sensible. En dicha definición se encuentran los datos relativos al estado civil de las personas, a su profesión u oficio y a su calidad de comerciante o de servidor público. De ese modo, estos pueden estar contenidos en registros públicos, documentos públicos, gacetas y boletines oficiales y sentencias judiciales debidamente ejecutoriadas que no estén sometidas a reserva, entre otros.

De ahí que la definición de dato sensible cobra mayor importancia (Decreto No. 1377, art. 3.3)), pues todo **aquel dato que sea sensible, por definición y naturaleza de este, no podrá ser público**. En ese sentido, para su tratamiento, se requerirá de una de las excepciones establecidas por la Ley No. 1581 (2012), entre las que se encuentran que se cuente con el consentimiento expreso y previo.

En cuanto a las excepciones al consentimiento, el artículo 10 de la Ley No. 1581 (2012) sostiene que en caso del tratamiento de **datos de naturaleza pública** no será vital la

autorización previa e informada del titular de datos personales[94]. En ese sentido, es relevante citar el artículo 2.2.2.25.2.2 del Decreto 1074, Decreto Único Reglamentario del Sector Comercio, Industria y Turismo (2015), que señala que *los datos personales que se encuentren en fuentes de acceso público pueden ser tratados siempre y cuando, por su naturaleza, sean datos públicos*[95]. Así, se fija el criterio de que, aunque un dato personal sea de acceso público, eso no significa que necesariamente sea un dato de naturaleza pública.

De otro lado, de acuerdo con la Ley No. 1266 (2008), Colombia cuenta dos autoridades diferentes que resguardan la protección de datos personales. La primera es la Superintendencia de Industria y Comercio (SIC), a través la Delegatura de la Dirección de Protección de Datos Personales, que, conforme con la Ley No. 1581 (2012), es la máxima autoridad con facultades para investigar e imponer sanciones que incumplan con los principios de tratamiento. La segunda es la Superintendencia Financiera de Colombia (SFC), que actúa como supervisor de las instituciones financieras y centrales de riesgo, siendo que verifica el cumplimiento de la Ley No. 1266 (2008).

Otro aspecto relevante es la obligación de informar del responsable o encargado de los datos personales, que se desarrolla en la Sentencia C-748 de la Corte Constitucional de Colombia (2012), la cual estableció que el responsable o el encargado del tratamiento deben comunicar, entre otros, el propósito del procesamiento de los datos personales.

En cuanto a jurisprudencia relevante en materia de *Web Scraping*, la SIC ordenó a LinkedIn Corporation y LinkedIn Ireland Unlimited Company (2023) a reforzar sus medidas de seguridad frente al *Web Scraping*, dado que exponían datos personales de más de 12 millones

[94] Artículo 10. Casos en que no es necesaria la autorización. La autorización del Titular no será necesaria cuando se trate de:
a) Información requerida por una entidad pública o administrativa en ejercicio de sus funciones legales o por orden judicial;
b) Datos de naturaleza pública;
c) Casos de urgencia médica o sanitaria;
d) Tratamiento de información autorizado por la ley para fines históricos, estadísticos o científicos;
e) Datos relacionados con el Registro Civil de las Personas.
Quien acceda a los datos personales sin que medie autorización previa deberá en todo caso cumplir con las disposiciones contenidas en la presente ley.

[95] Artículo 2.2.2.25.2.2. Autorización.
(...)
Los datos personales que se encuentren en fuentes de acceso público, con independencia del medio por el cuál se tenga acceso, entendiéndose por tales aquellos datos o bases de datos que se encuentren a disposición del público, pueden ser tratados por cualquier persona siempre y cuando, por su naturaleza, sean datos públicos.

de usuarios colombianos[96]. Es por ello que, mediante la Resolución No. 71406 de 2023, la SIC le ordenó que refuerce sus medidas de seguridad en un corto plazo de cuatro (04) meses.

Es fundamental mencionar que, bajo el ordenamiento jurídico de Colombia, la información personal que es "públicamente disponible" o "accesible al público" no es per se, información de "naturaleza pública". La Resolución No. 63771 (2022) contra Dataset Tecnologies S.A.S., empresa dedicada a operar como plataforma de antecedentes, denuncias, demandas y procesos judiciales que utilizaba *Web Scraping* para alimentar su plataforma, establece que **los datos contenidos en bases de datos o entidades públicas no son siempre de carácter público**, **sino que deben ser analizados para verificar la naturaleza real de los datos** y, considerando ello, sí es necesario la autorización de los titulares, de conformidad con la normativa de protección de datos personales.

Asimismo, la SIC ordenó a Dataset Technologies S.A.S a dejar de utilizar tecnologías automatizadas, incluyendo, pero no limitándose, al *Web Scraping* para la consulta de información de los titulares de las páginas de consulta de procesos judiciales a través de su portal www.datajuridica.com. Siendo esto así, indicó que debió obtener el consentimiento expreso, previo e informado de los titulares; y le ordenó eliminar todos los datos personales que había extraído, entre otros aspectos.

El análisis del legislador de la SIC consiste en uno prohibitivo en su totalidad, cuando en realidad pudo haber recogido un análisis de la pertinencia en la extracción de los datos para determinar si había proporcionalidad entre el dato recolectado, el mecanismo utilizado y la finalidad que se persigue.

Luego, se cuenta con la Resolución No. 58834 de 2023, nuevamente en contra de Dataset Technologies S.A.S., en donde la SIC mantuvo el mismo análisis que en su anterior pronunciamiento, indicando que el simple motivo de que un dato personal sea accesible públicamente entonces no es, como tal, de naturaleza pública. Asimismo, señaló que el *Web Scaping* es posible siempre y cuando se respete la normativa de protección de datos, incluyendo principalmente el cumplimiento del principio de veracidad, es decir, se recopile información que sea cierta o verídica (SIC, 2023).

[96] Se define como 'raspado' de información de sitios web a través de software, que en un alto porcentaje simulan la navegación de un humano

En ese sentido, recomendó implementar algunas medidas técnicas y organizativas para mitigar los riesgos del *Web Scraping*, como:

> (i) Asignar una estructura organizativa o roles determinados dentro de la empresa para poder identificar, implementar y monitorear controles de seguridad contra el *Web Scraping*.
> (ii) Limitar la "velocidad de acceso" de otros perfiles o cuentas a un número determinado de visitas por hora o día para evitar actividades sospechosas.
> (iii) Supervisar la rapidez y agresividad con la que una nueva cuenta comienza a buscar otros usuarios.
> (iii) Tomar medidas para detectar *scrapers* con el objetivo de identificar patrones de actividad de "bots".
> (iv) Incorporar CAPTCHAs para evitar que los extractores de datos sean "bots". De ser el caso, bloquear la dirección IP.
> (v) Si se sospecha o confirme el *Web Scraping*, tomar medidas legales como enviar comunicaciones de "cese y desistimiento", exigir la eliminación de la información, obtener confirmación de la eliminación y adoptar medidas legales para respetar los términos y condiciones respecto a la prohibición de extracción de datos personales.

Finalmente, podemos señalar que Colombia se ha pronunciado sobre el tratamiento de datos con el *Web Scraping* considerándolo válido siempre y cuando se cumpla con la normativa de protección de datos personales. Ello se ha evidenciado en los pronunciamientos con Dataset Technologies S.A.S., siendo uno de los pocos países de la región que ha tomado criterio y posición autónoma sobre la materia.

CAPÍTULO IV: IMPLICANCIAS DEL WEB SCRAPING EN LA REGULACIÓN PERUANA DE PROTECCIÓN DE DATOS PERSONALES

Dadas las regulaciones vigentes en materia de extracción de datos personales y sensibles, es evidente que el *Web Scraping* configura una técnica sumamente influyente y famosamente utilizada por distintos actores del medio. Considerando que los datos, en su mayoría, son extraídos de fuentes de acceso público, su extracción, recopilación y estructuración en una base de datos constituyen elementos mucho más valiosos que cada perfil por separado.

Al respecto, es importante realizar una elemental diferenciación. Existen los (i) datos extraídos de fuentes públicas, facilitados por los propios usuarios o disponibles en el Internet; y (ii) datos obtenidos a través de una filtración o fuga de datos mediante piratas informáticos o también denominados [97]. Generalmente se suele llamar a ambos supuestos -erróneamente - como filtración de datos (*data leaks or data break*); sin embargo, esta definición calzaría únicamente en el[98].

Por ello, el *Web Scraping* califica en el primer supuesto, es decir, la extracción de aquellos datos extraídos de páginas web (Kinsta, 2022) que califican como fuentes públicas de manera automatizada por medio de scripts o programas (Sinche y Torres, 2021; Mitchel, 2018) y que son facilitados en su gran mayoría por los mismos usuarios. Es cada vez más frecuente que dicha práctica se ejecute involucrando datos personales y sensibles en grandes cantidades, encontrándonos con altos e importantes problemas de privacidad. Conforme hemos podido evidenciar, distintos países extranjeros y vecinos de la región han identificado sus implicancias y efectos a nivel práctico y jurídico; sin embargo, aún se encuentra en un terreno de exploración y análisis, generando consultas al público con el propósito de recepcionar posturas y determinar si se debe aplicar (o no) un marco orientador para su aplicación conforme con la normativa de protección de datos personales según cada jurisdicción.

Al respecto, de la revisión de jurisprudencia y doctrina extranjera, podemos identificar que la técnica del *Web Scraping* puede utilizarse para distintos usos o propósitos, tales como los siguientes:

- Análisis de mercado: monitorear servicios o tendencias de diferentes sitios web de la competencia para establecer mecanismos y/o estrategias comerciales.
- Investigación académica: extraer información de distintas fuentes públicas en Internet para posterior análisis estadísticos e investigación.

[97] Actualmente, uno puede tomar conocimiento si es que ha sido víctima de *Web Scraping* a través del siguiente website: https://haveibeenpwned.com/.

[98] Actualmente, el Perú cuenta con la Ley de Delitos Informáticos, Ley No. 30096, en el cual se detallan una serie de delitos relacionados a las nuevas tecnologías. Dentro del artículo 2 se tipifica el delito de acceso ilícito, siendo que se sanciona con una pena privativa de libertad no menor de uno ni mayor de cuatro años y con treinta a noventa días-multa a todo aquel que deliberada e ilegítimamente accede a todo o en parte de un sistema informático, o se excede en lo autorizado.

- *E-commerce*: permite consolidar información comercial en un solo espacio para crear ofertas en un sitio funcional. Por ejemplo, las páginas web que agrupan precios de boletos de avión o buses.
- Marketing: Extraer comentarios, reacciones y reseñas de clientes y/o potenciales clientes desde plataformas de redes sociales y distintas páginas web para determinar las preferencias del consumidor.
- Entrenamiento de Inteligencia Artificial: la extracción de volúmenes de información para el entrenamiento de la tecnología *machine learning*.
- Interés público: extraer información para fines periodísticos en temas de relevancia pública.

Sin embargo, también existen finalidades maliciosas en donde se puede utilizar el *Web Scraping*, tales como:

- La elaboración de bases de datos de personas naturales para la ejecución de delitos informáticos, tales como el phishing o suplantación de identidad en medios digitales.
- La extracción de correos electrónicos para la remisión de publicidad sin que se haya obtenido el debido consentimiento.
- La extracción de contenidos protegidos por la propiedad intelectual para su posterior publicación sin el debido permiso o consentimiento.
- Sobrecargar servidores ocasionando una baja de un sitio objetivo.
- Extraer datos personales para el perfilamiento comercial de potenciales clientes sin el debido consentimiento, así como la extracción de información para monitorear las preferencias en línea y comercializarlas con terceros.

Es así como aterrizamos al territorio peruano, en donde la técnica del *Web Scraping* no deja de ser ajena a las tendencias internacionales. Esto considerando que aún con la falta de investigaciones o estudios realizados sobre la materia, es evidente que nos encontramos frente a una realidad latente, en cuyo caso tales extracciones de datos personales son utilizadas para distintas finalidades ilegales, siendo materia del presente trabajo la violación abrupta de los principios de la LPDP (2011) y su Reglamento (2013).

Es pertinente analizar lo que consideramos un tratamiento de datos personales conforme a lo dispuesto en nuestra legislación. Como se ha revisado en el Capítulo II, el tratamiento de datos personales consiste en cualquier operación o procedimiento técnico, sea automatizado o no,

que permita la (...), extracción (...) de datos personales (art. 2.19, 2011). Al respecto, el *Web Scraping* es una técnica de extracción de datos personales de sitios web, pues implica la automatización del proceso de extracción de datos utilizando scripts o programas desde el barrido de información disponible en el Internet para posteriormente utilizar y/o almacenar dichos datos.

A la fecha, no encontramos una definición establecida por parte de la ANPD de lo que se puede entender por extracción de datos personales al constituir ello un tipo de tratamiento. Sin embargo, consideramos que la extracción de datos personales reside en identificar y aislar datos personales y/o sensibles en un formato automatizado (o no) sin importar su estructura para una finalidad (o finalidades) específica(s). Una vez que se cuenta con la extracción de la data, se procede con la recopilación de los datos personales y/o sensibles, que consiste en la recolección, síntesis, organización (ordenamiento o clasificación) y comprensión de los datos extraídos.

Por ello, identificamos al *Web Scraping* como una las técnicas que constituyen el primer escalón o paso dentro del tratamiento de datos personales, pues una vez que ocurre la extracción de la data se genera la recopilación, utilización u organización de los mismos.

Al amparo de la naturaleza y funcionamiento del *Web Scraping*, en conjunto con la LPDP (2011) y su Reglamento (2013), podemos concluir que dicha técnica es factible de ejecución siempre y cuando se cuente con el consentimiento expreso, informado, previo e inequívoco del titular de datos personales. No obstante, tal escenario no es recurrente en la práctica - por no decir nula -, pues generalmente se extraen datos personales y sensibles de aquella información accesible y/o disponible en el Internet. En consecuencia, ¿cómo realizar el tratamiento de datos personales mediante la extracción de los mismos vía *Web Scraping* de no contar con el consentimiento del titular? La citada pregunta será respondida en el siguiente acápite.

4.1 *Web Scraping* y su aplicabilidad en las fuentes de acceso al público

De acuerdo a lo observado en el Capítulo II del presente trabajo, el principio del consentimiento implica que será lícito si este ha sido libre, previo, expreso, informado e inequívoco. No obstante, existen excepciones a dicho consentimiento, siendo uno de ellos **cuando sean datos contenidos o destinados a ser fuentes accesibles para el público** (inciso 11, artículo 2 de la

LPDP (2011)). Para nuestro legislador, las "*fuentes accesibles para el público*" son aquellos **bancos de datos personales de la administración pública o privada** que **pueden ser consultados por cualquier persona**, realizando un abono previo, de ser el caso, siendo que el Reglamento determina las fuentes de acceso para el público en una lista taxativa (artículo 2. 11, LPDP, 2011)

Es así que, nuestro legislador, entiende como **fuentes de acceso para el público** como aquel que (i) es de acceso libre (gratuito) o (ii) se encuentre sujeto a un abono. Estas dos clasificaciones de fuentes de acceso se encuentran inmersas en los ocho (08) tipos detallados en el artículo 17 del Reglamento que se explicarán más adelante. Sin embargo, tales fuentes de acceso al público deberán respetar los principios de la LPDP (2011) y su Reglamento (2013), tales como el principio del consentimiento, finalidad, calidad y seguridad.

Ante lo señalado, no debemos confundir lo que entiende por "*fuentes accesibles para el público*" por "*información pública*", en cuyo caso este última es aquella información (documentos, fotografías, grabaciones), bajo cualquier formato, que haya sido creada u obtenida por la entidad pública o se encuentre bajo su control (Artículo 10 del Texto Único Ordenado de la Ley No. 27806, Ley de Transparencia y Acceso a la Información Pública). Así, la **información pública que deba ser entregada es un tipo de fuente accesible para el público** (inciso 8, artículo 17 del Reglamento), por lo que no son conceptos equivalentes o análogos al regirse por otro cuerpo normativo, es decir, por el Texto Único Ordenado de la Ley No. 27806, Ley de Transparencia y Acceso a la Información Pública (2017) ("en adelante, "LTAIP[99]".

Es importante señalar que **no toda información pública es un tipo de fuente accesible para el público**, pues, de acuerdo con la LTAIP (2017), la información pública que deba ser entregada no debe estar inmersa en las excepciones restrictivas[100] de la LTAIP (2017). Asimismo, cada solicitud de información pública constituye un caso aislado y único que se analiza acorde a las circunstancias y a la LTAIP (2017). En ese sentido, **únicamente la**

[99] La DGTAIPD, según el artículo 70 del Decreto Supremo 013-2017-JUS, dirige a la Autoridad Nacional de Transparencia y Acceso a la Información Pública (ANTAIP) y a la ANPD.

[100] Tales como la información secreta, reservada y confidencial, conforme con los artículos 15, 16 y 17 de la LTAIP. Estas excepciones deberán ser interpretadas de forma restrictiva y debidamente fundamentadas (Sentencia recaída en el Exp. No. 01593-2021-PHD/TC). Una excepción restrictiva resaltante es el que incluye a los datos personales cuya publicidad ocasione una vulneración a la intimidad personal y familiar (artículo 15 de la LTAIP).

información pública susceptible a ser entregada constituirá un tipo de fuente accesible para el público.

El Tribunal Constitucional entiende que toda persona debe gozar de una legítima expectativa de protección y respeto por su vida privada, en tanto el dato creado sobre la misma le permitirá desarrollarse (Tribunal Constitucional, 2005.. Por ello, la información pública podrá ser fuente de acceso al público siempre y cuando se permita en función a las circunstancias y a la naturaleza de la información, es decir, no nos encontremos inmersos en una excepción restrictiva.

Una vez analizado lo dispuesto en la Ley, resaltamos lo que indica nuestro Reglamento en el artículo 17, en donde se detallan expresa y taxativamente los ocho (08) tipos de fuentes acceso al público. Estos incisos deberán interpretarse y aplicarse en el marco de los principios de la LPDP (2011) y el Reglamento (2013):

1) "Los medios de comunicación electrónica, óptica y de otra tecnología, si el lugar que contiene los datos personales facilita la información al público y que pueda consultarse abiertamente."

> Si bien no se cuenta con pronunciamientos de la ANPD y del Tribunal Constitucional de lo que podemos entender por "*medios de comunicación electrónica, óptica y de otra tecnología*", resaltamos a los medios de comunicación electrónica, los cuales son aquellas herramientas que permiten la transmisión y recepción de información electrónicas, tales como las aplicaciones móviles, plataformas de *streaming*, **sitios web**, radio y prensa por internet, televisión por cable/satélite electrónica y los lenguajes de programación. Es mediante los medios de comunicación electrónica en donde se puede ejecutar el *Web scraping.*
>
> Por otro lado, los medios de comunicación óptica constituyen aquellas herramientas que se basan en transmisiones pulsadas por energía mediante cables de fibra óptica, mientras que aquellos que pertenecen a "*otra tecnología*" son aquellos que no pertenecen a ninguna de las dos definiciones anteriormente expuestas (por ejemplo, las tecnologías de red).

En cuanto al caso particular del Internet, la ANPD se ha pronunciado y ha determinado que en sí misma no es una fuente de acceso al público, sino un canal o puente por el cual se puede acceder a fuentes de información (Oficio No. 156-2017-JUS/DGTAIPD). Esto porque dentro de internet existe información pública y privada, por lo que la condición de fuente de acceso pública dependerá de las condiciones de privacidad del sitio web y con lo dispuesto por la LPDP (2011). En consecuencia, la información contenida en Internet y la nube quedará protegida y tratada conforme a la LPDP (2011) y el Reglamento (2013).

2) Las guías telefónicas, independientemente del soporte donde se encuentren y en los términos de su regulación específica.

Este numeral se refiere a las guías telefónicas que almacenan datos personales de usuarios de servicios de telefonía fija, sin importar (i) el soporte, pudiendo ser físicos (páginas blancas) o electrónicos (páginas web), y (ii) considerando la regulación sectorial aplicable, en este caso el Texto Único Ordenado de las Condiciones de Uso de los Servicios Públicos de Telecomunicaciones, aprobado mediante Resolución de Consejo Directivo No. 138-2012-CD/Osiptel, y sus normas conexas.

Tales guías telefónicas contienen, como mínimo, el nombre del abonado (nombres y apellidos), dirección de instalación y número de la línea telefónica que le ha sido asignada, en los cuales aparecerán listadas de forma gratuita dentro de una guía telefónica (Resolución Directoral No. 1085-2022-JUS/DGTAIPD-DPDP, 2021). Cualquier otra información adicional que se quiera introducir en la guía telefónica requerirá el consentimiento expreso, previo, inequívoco e informado del titular.

Asimismo, conforme indica el artículo 101 de la citada Resolución de Consejo Directivo No. 138-2012-CD/Osiptel, los titulares de datos personales pueden solicitar su exclusión de la referida guía, es decir, ejercer su derecho de cancelación (2012). Tal derecho incluye el servicio de información de guía telefónica y a la información incluida en la página web de las operadoras.

Finalmente, si bien es cierto que no se requiere el consentimiento del titular para el tratamiento de datos personales, se deberá cumplir con los demás principios de la LPDP

(2011) y su Reglamento (2013), como es el derecho de información desarrollado en el artículo 18 de la LPDP (2011) (Resolución No. 1085-2022-JUS/DGTAIPD-DPDP).

3) Los diarios y revistas independientemente del soporte donde se encuentren y en los términos de su regulación específica.

Siguiendo con el numeral 3 que corresponde a los diarios y revistas, nos encontramos frente a los medios físicos tradicionales también conocidos como la prensa por escrito. Dentro de esta categoría se incluye a las gacetas y los boletines oficiales.

4) Los medios de comunicación social

Si bien la ANPD no ha desarrollado a cabalidad dicho concepto, entendemos como medios de comunicación social a todos aquellos que permiten la interacción, intercambio de información y creación de contenido por parte de los usuarios. Incluyen programas de televisión, portales de noticias (físicos o digitales), radio, prensa, diarios, revistas y blogs (físicos o digitales), foros, comunidades en línea, y artículos de opinión, calificados como medios masivos. Es importante destacar que, en el contexto digital, dicha categoría agrupa a las **redes sociales**, pues se trata de plataformas de comunicación que permiten que los individuos creen redes de usuarios que comparten intereses comunes (pp. 4-5, Dictamen 5/2009 del Grupo de Trabajo del Artículo 29, 2009). Estos cuentan con las siguientes características: (i) los usuarios deben proporcionar datos personales para generar su perfil, (ii) las redes sociales proporcionan herramientas que permiten a los usuarios contar con su propio contenido en línea, y (iii) operan gracias a los contactos que genera el propio usuario.

De acuerdo con nuestro Tribunal Constitucional, las redes sociales son medios de comunicación (Sentencia recaída en el Exp. 00442-2017-PA/TC) y promueven el entretenimiento (Sentencia recaída en el Exp. 01163-2022-PHD/TC), siendo que permiten que las personas compartan información mediante la creación de sus perfiles. Asimismo, se diferencian de un medio tradicional al depender de la interacción de las personas para generar contenido dentro del espacio, utilizando a la tecnología como conductor de la información (Banco Interamericano de Desarrollo, s.f.).

A la fecha, la ANPD ha emitido posturas distintas en el marco del tratamiento de datos personales dentro de las redes sociales por parte de un tercero. En el 2014 sostuvo que se requerirá del consentimiento del usuario dentro de la red social para que se pueda configurar su tratamiento por parte de un tercero, siendo que el usuario de la red social sea una persona natural (Oficio No. 569-2014-JUS/DGPDP).

Por otro lado, en el 2019, se desarrolló la postura en que no se requerirá del consentimiento del titular de datos personales cuando se trate de un perfil público (Resolución Directoral No. 1623-2019-JUS/DGTAIPD-DPDP, 2019), siempre que existan datos abiertos y disponibles. Sin embargo, ello no implica un tratamiento indiscriminado de los datos personales, sino que deben cumplir con los demás principios de la LPDP (2011) y el Reglamento (2013), tales como el principio de finalidad y calidad. Por ello, si el usuario incluye en la configuración de su red social cierta información personal de manera pública (por ejemplo, nombres completos e imagen), tales datos personales siguen perteneciendo a la esfera de la privacidad del usuario, por lo que su tratamiento deberá cumplir con la LPDP (2011) y su Reglamento (2013).

Al respecto, consideramos que no se deberá requerir del consentimiento del usuario de la red social, sea perfil público o privado, siendo en este último caso en la medida en que la información del usuario se encuentre disponible o abierta para consulta de cualquier persona[101], sin perjuicio del cumplimiento de los principios de la LPDP (2011) y el Reglamento (2013), como lo son el principio de finalidad y el de calidad.

5. Las listas de personas que forman parte de los grupos profesionales que incluyan solo datos de nombre, título, profesión, actividad, grado académico, dirección postal, número telefónico, número de fax, dirección de correo electrónico y quienes establezcan su pertenencia al grupo. Asimismo, en el caso de colegios profesionales, se podrán indicar, datos de sus miembros en particular como: número de colegiatura, fecha de incorporación y situación gremial en relación al ejercicio profesional.

[101] Por ejemplo, si bien el perfil es privado, de igual forma se puede consultar o es accesible al público sus nombres y apellidos, foto de perfil, lugar de residencia, entre otros. Todo lo anterior dependerá de los límites de privacidad que haya limitado el usuario, conforme cada caso concreto.

Por su parte, la ANPD se ha pronunciado sobre el presente numeral, advirtiendo que la finalidad es identificar a un profesional determinado, verificar su estatus de habilitación en el colegio profesional respectivo y contactarlo de ser necesario (Opinión Consultiva No. 749-2018-DGTAIPD). Esto siempre cumpliendo con el principio de finalidad, por lo que, por ejemplo, la publicación de datos personales del profesional de la salud como: colegio profesional, número de colegiatura, especialidad y turno/programación, son datos de acceso público siempre que el tratamiento tenga por objeto identificar a dichos profesionales de los grupos y colegios profesionales, en este caso de los servicios de la salud.

Asimismo, el principio de proporcionalidad busca evaluar si los datos proporcionados son necesarios y suficientes para ejercer las funciones del servicio frente a los usuarios. Así, sobre el tratamiento del número del DNI de los médicos que atienden en centro profesionales de la salud, este no resulta necesario para identificar al médico y conocer si se encuentra habilitado para el ejercicio de la profesión médica (Opinión Consultiva No. 020-2023-DGTAIPD).

6. Los repertorios de jurisprudencia, debidamente anonimizados.

El presente numeral hace referencia a los repertorios de jurisprudencia, incluyendo a los digitales (por ejemplo, las sentencias publicadas en el sitio web del Poder Judicial), las cuales deberán estar anonimizados. La anonimización de los datos se define, en el artículo 14.2 de la LPDP (2011), como el tratamiento de datos personales que no permite la identificación o no hace identificable al dueño de los datos, siendo que la medida será irreversible.

Por otro lado, la calidad de los repertorios de jurisprudencia se sostiene en el artículo 139.4 de la Constitución, que constituye un principio y derecho de la función jurisdiccional la publicidad de los procesos, salvo que la ley disponga lo contrario. Al respecto, la ANPD se ha pronunciado indicando que la publicación de las sentencias y resoluciones judiciales responden a un innegable interés público de conocimiento de los criterios jurisprudenciales, además de la fundamentación de las decisiones del legislador que sirven para futuras interpretaciones. Es así que, resulta intrascendente conocer los datos personales de las partes, pues la materia relevante de análisis recae

en el contenido jurídico de las sentencias (Resolución No. 3551-2018-JUS/DGTAIPD-DPDP, foja 35-38). Asimismo, la ANPD extiende dicho criterio para resoluciones administrativas (Resolución No. 1264-2021-JUS/DGTAIPD-DPDP).

En cuanto a la anonimización, el fundamento del mismo se encuentra contenido dentro del principio de proporcionalidad regulado en el artículo 7 de la LPDP (2011), que dispone que todo tratamiento debe ser adecuado, relevante y no excesivo para la finalidad para la cual fueron recopilados los datos personales. De ese modo, la ANPD enfatiza la relevancia de la anonimización, incluyendo que cualquier persona (natural o jurídica) que comparta jurisprudencia sin cumplir con anonimizar los datos personales de las partes, será responsable por el tratamiento de los mismos (Resolución No. 3442-2021-JUS/DGTAIPD-DPDP).

7. Los Registros Públicos administrados por la Superintendencia Nacional de Registros Públicos - SUNARP, incluyendo cualquier otro registro o banco de datos calificado como público, en virtud de la ley.

Respecto al presente acápite, se incluye a todos aquellos registros públicos de la SUNARP y a cualquier otro banco de datos de naturaleza pública. Al respecto, es preciso indicar que no todos los bancos de datos de propiedad del Estado son catalogados como públicos. Por ejemplo, en una reciente sentencia del Tribunal Constitucional, se advirtió que la base de datos personales del Sistema de Registro de Denuncias –SIDPOL de la Dirección de Criminalística de la Policía Nacional del Perú no constituía una base de datos de carácter público ni es susceptible de acceso a terceros (Sentencia recaída en el Expediente No. 02839-2021-PHD/TC, foja 13).

De otro lado, el Tribunal Constitucional también se ha pronunciado para ratificar ciertas bases de datos susceptibles de acceso al público, como es el caso de la base de datos de registro de requisitorias (Sentencia recaída en el Expediente No. 5060-2009-PHD/TC).

8. Las entidades de la Administración Pública, en relación a la información que deba ser entregada en aplicación de la Ley Nº 27806, LTAIP.

El presente acápite se refiere a lo advertido en párrafos anteriores en cuanto a la naturaleza de la información pública en contraposición de lo que contiene la definición *fuentes accesibles para el público*. Al respecto, toda documentación que una entidad posea, administre o haya generado mediante el ejercicio de sus funciones, sin perjuicio del origen, medio de almacenamiento, o utilización, constituirá información de naturaleza pública (Resolución recaída en el Expediente No. 01912-2021-JUS/TTAIP). El concepto de información pública se basa en el artículo 2 inciso 5 de la Constitución y ninguna entidad estatal o persona de derecho público puede exceptuarse de la obligación de otorgar aquella información de naturaleza pública (Sentencia recaída en el Expediente 00937-2013-PHD/TC). Asimismo, el artículo 43 del Reglamento de la LTAIPD, aprobado mediante Decreto Supremo No. 007-2024-JUS, ("Reglamento de la LTAIPD") establece el principio de *transparencia proactiva*, en el que las instituciones públicas deberán publicar toda información adicional que resulte útil y oportuna para los ciudadanos, sin perjuicio de lo requerido por la LTAIPD u otra normativa aplicable.

De lo señalado, es preciso indicar que no todos los presuntos "*datos personales*" en posesión de entidades públicas se encuentran inmersos en la Ley y Reglamento. Por ejemplo, el Tribunal de Transparencia y Acceso a la Información Pública - TTAIP señaló que los números de celulares adquiridos y otorgados a los funcionarios de una entidad estatal son carácter público, exceptuándose la posibilidad de ser datos personales (Sentencia recaída en el Expediente No. 00673-2020-JUS/TTAIP).

Asimismo, los correos electrónicos institucionales entregados a los servidores o funcionarios públicos son de carácter público, pues no son utilizados para actividades privadas (Sentencia recaída en el Exp. No. 04792-2017-PHD/TC). De hecho, se cuenta con la Directiva No. 005-2003-INEI/DTN, "Normas para el uso del servicio de correo electrónico en las entidades de la Administración Pública", en donde los correos electrónicos de los funcionarios y servidores públicos deberán ser usados para el cumplimiento de sus funciones (Artículo 5.2, Directiva No. 005-2003-INEI/DTN). En esa línea, la Exposición de Motivos del Decreto Supremo 011-2018-JUS resalta la naturaleza pública de los correos electrónicos de los funcionarios y servidores del Estado al encontrarse en el principio de publicidad y máxima divulgación. De hecho, nuestro Tribunal Constitucional establece que la publicidad constituye la regla general

dentro de las actividades públicas, mientras que el secreto será la excepción (Sentencia recaída en el Expediente No, 2579-2003-HD/TC).

En cuanto al contenido de la información de los correos electrónicos de los servidores y funcionarios públicos, estos serán de naturaleza pública, siempre que no nos encontremos frente a las excepciones restrictivas. (Opinión Consultiva 034-2022-JUS/DGTAIPD; Sentencia recaída en el Exp. 04792-2017-PHD/TC). Ello en razón a que los correos electrónicos de la entidad han sido creados para permitir la ejecución de sus labores y actividades, considerando también que representan un costo a cargo de la entidad pública. De acuerdo con el artículo 36.2 del Reglamento de la LTAIPD, será el funcionario o servidor quien pondrá a disposición la información solicitada al área encargada interna de la propia entidad pública - como la Oficina de Tecnologías de la Información - (Opinión Consultiva No. 51-2018-DGTAIPDP) y, de ser el caso, deberá fundamentar debidamente el motivo por el cual no otorga información que considera parte de las excepciones restrictivas, es decir, una *motivación cualificada* (Sentencia recaída en el Expediente 03035-2012-PHD/TC).

En el caso de la información contenida en correos electrónicos de exfuncionarios o exservidores públicos, también representan de acceso público siempre y cuando no nos encontremos dentro de las excepciones restrictivas. Al respecto, el artículo 37 del Reglamento de la LTAIPD sostiene que cuando el personal público se desvincula de entidad, este deberá disponer toda la información de su correo electrónico al ente estatal, quien será quien determine la accesibilidad (o no) de la información contenida en el correo electrónico institucional.

Sin embargo, los datos de contacto como domicilio, correo y teléfonos personales de los funcionarios o servidores públicos no constituyen información pública (Expediente No. 01912-2021-JUS/TTAIP).

Por otro lado, para nuestro TTAIP (Opinión Técnica Vinculante No. 000001-2021-JUS/TTAIP-SP), cuando se presente una solicitud de acceso a la información pública de uno mismo y que esté relacionado a sus datos personales, entonces deberá aplicarse la Ley y su Reglamento, dejando de pertenecer a la esfera de información pública. En esa misma línea advierte nuestro Tribunal Constitucional (Sentencia recaída en el

Expediente 7189-2013- PHD/TC) al señalar que cuando se involucran datos personales en una solicitud de acceso a la información sobre sí mismo, entonces se trata de un ejercicio del derecho a la autodeterminación informativa y no al de acceso a la información pública. Por ello, para nuestro legislador, una vez que se ingresa a la aplicación de la Ley y Reglamento, entonces tal solicitud se regirá por los principios de tal normativa, es decir, por el principio de finalidad, calidad, proporcionalidad y seguridad.

Finalmente, advertimos que no toda información pública es una *fuente de acceso para el público*, pues para serlo deberá estar fuera de las excepciones restrictivas tipificadas en la LTAIP y deberá analizarse conforme a las circunstancias de cada caso concreto.

Dada la explicación de cada tipo de fuente de acceso para el público, identificamos que todos ellos son susceptibles a la ejecución de la técnica del *Web* Scraping, siendo las fuentes de acceso más usadas (i) los medios de comunicación social (artículo 17.4 del Reglamento), específicamente por la extracción de información por parte de las plataformas redes sociales, y (ii) las plataformas de comercio electrónico (artículo 17.1 del Reglamento).

En caso se tercerice la extracción de datos personales obtenida de fuentes accesibles al público para una finalidad comercial (compraventa), tanto el contratante como el contratista deberán obtener el consentimiento de los titulares, de ser aplicable, siendo que cada uno de ellos será titular de banco de datos personales al determinar el tratamiento para distintas finalidades (Informe No. 06-2017-JUS/DGTAIPD-DPDP).

Si bien los ocho (08) tipos mencionados previamente en nuestro Reglamento no requieren del consentimiento del titular de datos para su tratamiento, ello no exime del respeto y cumplimiento de los principios de la LPDP (2011) y el Reglamento. Ahora bien, ¿qué implica el cumplimiento de los citados principios? La ANPD, junto con la doctrina autorizada, se ha pronunciado sobre el particular en cuanto al tratamiento de datos obtenidos mediante fuentes de acceso al público.

Tabla 7.
Principios de la LPDP (2011) sobre fuentes de acceso al público

Principio	Contenido

Finalidad	Implica que los datos personales deben ser extraídos para una finalidad determinada, lícita y explícita, es decir, el tratamiento no puede excederse a otra finalidad (artículo 6 de la LPDP (2011)). Ello involucra que la finalidad deba ser explícita, clara y no admite ningún tipo de confusiones (artículo 8 del Reglamento). La ANPD ha señalado que los datos contenidos en fuentes de acceso al público deberán utilizarse para las finalidades creadas y por las que han sido puestas a disposición del público (Oficio No. 749-2018-JUS/DGTAIPD). En caso de requerir tratamientos para finalidades distintas, se deberá contar con el consentimiento del titular del dato personal (Informe No. 06-2017-JUS/DGTAIPD-DPDP). Por ejemplo, si se obtienen datos personales para la evaluación de un préstamo hipotecario mediante fuentes de acceso público, entonces tales datos no podrán remitirse a terceros para ofrecer servicios publicitarios (Oficio No. 749-2018-JUS/DGTAIPD).
Proporcionalidad	El tratamiento de datos personales deberá ser relevante, adecuado y no excesivo a la finalidad que motiva su tratamiento (artículo 7 de la LPDP (2011)). Los datos deben estar relacionados con el fin, es decir, deben ser usados para los fines que fueron registrados sin afectar a los sujetos (Quiroga, 2013). Por ello, en caso de datos personales obtenidos por fuentes de acceso público, tal tratamiento no deberá ser excesivo de la propia finalidad publicada al público, siendo que su tratamiento se encuentra limitado por la misma finalidad que motivó su publicación.
Calidad	Los datos personales deben responder a la realidad con precisión, considerando la presunción de que los datos personales proporcionados directamente por el titular son exactos (artículo 9 del Reglamento). Así, los mismos deberán ser veraces, actualizados, pertinentes y necesarios a la finalidad que motivó su extracción. Lo anterior incluye que su conservación sea conforme con las medidas de seguridad pertinentes y por el tiempo necesario (artículo 8 de la LPDP (2011)). Cavero y Holguín (2013) sostienen que el principio de calidad busca el uso eficiente de la información. Por ello, los datos accesibles al público son veraces, actualizados y pertinentes en la medida de la finalidad que motivó su publicación.
Seguridad	El titular de banco de datos personales y el encargado de tratamiento deberán adoptar las medidas organizativas, legales y técnicas necesarias. Estas medidas deberán ser apropiadas al tratamiento a ejecutar y con la categoría de datos aplicable (artículo 9 de la LPDP (2011)). Ello incluye la protección a la adulteración, pérdida o desviaciones de información, sea de acción humana o medios técnicas (artículo 10 del Reglamento).
Nivel de protección adecuado	Se debe garantizar un nivel suficiente de protección para los datos personales que se vayan a tratar o, al menos, equiparable a lo

	previsto por la LPDP (2011) dentro del flujo transfronterizo (artículo 11 de la LPDP (2011)).
Legalidad	Se prohíbe la recopilación (y extracción) de los datos personales por medios fraudulentos, desleales o ilícitos. Tal como señala Cavero y Holguín (2013), este es el principio rector en cualquier tipo de normativa y significa el respetoal ordenamiento.
Disposición de recurso	El titular de datos personales cuenta con vías administrativas o jurisdiccionales necesarias para reclamar y hacer valer sus derechos ante vulneraciones (artículo 10 de la LPDP (2011)). En caso no se realice el tratamiento de datos obtenidos por fuentes de acceso al público, corresponderá que los titulares de datos inicien

Fuente: Elaboración propia

Es importante señalar que los citados principios son de carácter enunciativo y constituyen parámetros para la interpretación en la resolución de casos, así como completar vacíos en la legislación de protección de datos personales (artículo 12 de la LPDP (2011)).

Habiendo revisado la jurisprudencia aterrizada al tratamiento de datos personales obtenidos por fuentes de acceso público, evidenciamos que su tratamiento quedará limitado, principalmente, por el principio de finalidad, dado que se cuenta con el acceso a tales datos por la finalidad (o finalidades) que motivó (o motivaron) su publicación. A ello se incluye que tales datos serán veraces, proporcionales y contarán con las medidas de seguridad adecuadas por la finalidad que se pretendió perseguir y que, por tanto, permite su disponibilidad y consulta al público. Por ello, si es que se realiza un tratamiento distinto a dicha finalidad, entonces será necesario obtener el consentimiento previo, expreso, informado e inequívoco del titular.

4.2 Nuestra propuesta: hacia un adecuado tratamiento de datos personales

El Tribunal Constitucional ha reconocido que la tecnología permite la reproducción de datos de todo tipo a través de sistemas informáticos, motores de búsqueda y cualquier dispositivo tecnológico, haciendo que se logre una hipervisibilización de los mismos (Sentencia recaída en el Expediente No. 03041-2021-PHD/TC, foja 10). Debido al incontrolable y constante tratamiento de los datos personales, estos se han convertido en un valor económico al impulsar la economía tecnológica. Considerada como la moneda del nuevo milenio, el valor monetario de los datos personales es grande y sigue creciendo, [y por eso] distintos responsables de tratamiento sacan provecho de tal tendencia (Schwartz, 2004), sin considerar los principios de protección de datos personales inmersos en dicho tratamiento.

Como mencionamos anteriormente, Perú no es ajeno a dicha realidad y nuestra regulación no puede desconocer que la gran mayoría de datos personales extraídos por la técnica de *Web Scraping*, así como otras técnicas existentes, se realizan a través de fuentes de acceso al público. Lo anterior se encuentra demostrado a través de la diversa casuística extranjera, incluyendo que se ha suscrito un acuerdo internacional con el propósito de establecer medidas confrontativas frente al tratamiento de datos obtenidos por fuentes de acceso al público mediante el *Web Scraping* (Information Commissioner's Office, 2023). En consecuencia, remitiéndonos a nuestra hipótesis si es que se debe regular, *per se*, la técnica del *Web Scraping*, la respuesta resulta, a todas luces, negativa.

Al amparo de las nuevas tecnologías emergentes y la evolución tecnológica, tajantemente reconocido por nuestro Tribunal Constitucional, imponer regulaciones sobre la misma ralentiza la innovación y creación de nuevos productos y servicios que operan con *Web Scraping*, por lo que atenta contra su propia naturaleza. Asimismo, instaurar medidas que prohíban y eviten la ejecución de la técnica del *Web Scraping* en su totalidad crea barreras innecesarias para el funcionamiento de distintas tecnologías que requieren de la misma y no involucran, necesariamente, el tratamiento de datos personales.

En este caso, dado que el *Web Scraping* constituye un tipo de tratamiento al extraer datos personales, entonces le será aplicable la LPDP (2011) y su Reglamento. Dicho lo anterior, regular exclusivamente tal práctica resulta altamente ineficiente y redundante, pues actualmente existen principios y obligaciones en materia de la LPDP (2011) y su Reglamento que el *scraper* debe cumplir al momento de extraer los datos personales.

Recordemos que la finalidad de nuestro legislador es evitar la duplicidad normativa, siendo que será necesario una regulación solo si, además de cumplir con otros parámetros, contribuye a resolver o reducir un problema público en base a evidencia y que no cuenta con una protección jurídica como tal[102]. Así, se evidencia que el propósito de nuestro legislador y autoridades administrativas es evitar escenarios de regulaciones redundantes e injustificadas.

[102] Decreto Legislativo No. 1565
Artículo 4.- Principios de la mejora de la calidad regulatoria La mejora de la calidad regulatoria se sustenta, fundamentalmente, en los siguientes principios, sin perjuicio de la aplicación de otros que puedan ser establecidos en el Reglamento del presente Decreto Legislativo:
f. Necesidad: La regulación cuenta con evidencia previa que demuestre que la alternativa seleccionada sea la más beneficiosa. Esta contribuye con el objetivo de resolver, reducir los riesgos o mitigar un problema público identificado en base a evidencia.

Conforme a lo analizado en distintas regulaciones extranjeras y de la región, los responsables de tratamiento que ejecutan la técnica del *Web Scraping* deberán cumplir con la regulación vigente sobre la materia, es decir, con los principios y obligaciones tipificadas en la LPDP (2011) y su Reglamento. Aunado a ello, la ANPD será actor clave para el cumplimiento de tal normativa, pues en caso de incumplimiento, deberá aplicar las sanciones administrativas correspondientes.

No obstante, en el marco de la clara negligencia y vulneración de los principios en el tratamiento de datos personales (y sensibles) a través de la extracción de datos (como el *Web Scraping*) provenientes de las distintas páginas web o plataformas digitales alojadas en el Internet, es que se resalta la necesidad de revestir con mayor seguridad jurídica los principios que protegen a los titulares de datos personales, es decir, aquellos tipificados en la LPDP (2011) y Reglamento. Por ello, se propone una modificación parcial al artículo 14 de la LPDP (2011) bajo la siguiente fórmula normativa:

Tabla 8.
Propuesta normativa al amparo del artículo 14.2 de la LPDP (2011)

LPDP (2011)	**Propuesta normativa**
Artículo 14. Limitaciones al consentimiento para el tratamiento de datos personales No se requiere el consentimiento del titular de datos personales, para los efectos de su tratamiento, en los siguientes casos: 2. Cuando se trate de datos personales contenidos o destinados a ser contenidos en fuentes accesibles para el público. (...)	Artículo 14. Limitaciones al consentimiento para el tratamiento de datos personales No se requiere el consentimiento del titular de datos personales, para los efectos de su tratamiento, en los siguientes casos: 2. Cuando se trate de datos personales contenidos o destinados a ser contenidos en fuentes accesibles para el público, ***siempre y cuando se realice y documente, de manera previa al tratamiento, una evaluación del cumplimiento de los principios de la presente ley, conforme cada caso concreto.*** (Resaltado es nuestro) (...)

Fuente: Elaboración propia

En el proceso de generación de evidencia y diseño de la regulación, se analizan todas las alternativas racionales con la finalidad de elegir la mejor opción regulatoria.

A la fecha, nuestro Reglamento de la Ley 29733 plantea en su artículo 17 que el tratamiento de datos personales obtenidos por fuentes de acceso al público deberá ceñirse a los principios establecidos en la LPDP (2011) y su Reglamento. No obstante, la propuesta indica que dicho reconocimiento se realice dentro de la LPDP (2011), lo que constituye un hito elemental para que los responsables de tratamiento prioricen la revisión de su cumplimiento. Esto en tanto que, conforme los artículos 51 y 200.4 de la Constitución, el principio de jerarquía normativa radica en que una norma será jerárquicamente superior a otra cuando la validez de la segunda depende de la primera (Sentencia recaída en el Exp. 047-2004-AI/TC). En este caso, el Reglamento depende de la LPDP (2011), y esta última de la Constitución (artículo 2.6 de la Constitución).

Recordemos que los principios permiten la interpretación de las normas y son catalogados como mandatos de optimización que ordenan que algo sea ejecutado de acuerdo con el caso en concreto y en función a las posibilidades jurídicas existentes (Alexy, 2002). Aterrizando dicho concepto dentro de nuestra normativa de datos personales, los principios rectores orientan y fijan el comportamiento de los actores que participaran en el proceso de tratamiento. Es por ello que tal reconocimiento de carácter legal asegura un cumplimiento obligatorio e imperativo de los principios frente al tratamiento de datos obtenidos por fuentes de acceso al público, específicamente al principio de finalidad.

Por otro lado, tal evaluación deberá ser *ex ante*, pues, dada las condiciones y matices del caso en que nos encontremos, podremos dilucidar si el tratamiento perseguido a través de la extracción mediante el *Web Scraping* es legal. Es evidente que antes de ejecutar algún tipo de tratamiento se requiere, *prima facie*, verificar que se estén cumpliendo los principios rectores que determinan a todo tipo de tratamiento existente y que se realice una interpretación en cada caso concreto.

Asimismo, es vital que el análisis del cumplimiento de los principios de la LPDP (2011) y su Reglamento se encuentren documentados con el propósito que, en caso el titular del dato consulte cómo se han obtenido sus datos personales para la ejecución del tratamiento, entonces se cuente con una respuesta clara, completa y debidamente fundamentada por parte del responsable de tratamiento. En materia de seguridad de la información, documentar el tratamiento de datos personales es una práctica organizativa diligente y alineada con la

Directiva de Seguridad de la ANPD, además de constituir una práctica de Evaluación de Impacto del tratamiento de datos personales[103].

Esta medida también se encuentra complementada con el principio de responsabilidad proactiva o *accountability*, que consiste en la capacidad de demostrar que se está realizando el tratamiento de datos personales en función a los principios de la LPDP (2011) y su Reglamento. Este principio se basa en el equilibrio de la ética y responsabilidad directa de los responsables de tratar datos personales, considerando que la privacidad proactiva permite mitigar riesgos en materia de privacidad. Este principio de carácter preventivo fomenta la seguridad en todos sus parámetros y manifiesta un respeto por la privacidad de los titulares. Actualmente se encuentra en discusión su inclusión dentro de nuestra legislación mediante el Proyecto de Reglamento (art. VII, 2023)[104], aprobado mediante Resolución Ministerial No. 0270-2023-JUS ("Proyecto de Reglamento").

Habiendo señalado lo anterior, la propuesta de modificación normativa es imperativo y pertinente para garantizar un adecuado tratamiento de datos personales obtenidos por fuentes de acceso al público, en este caso, a través de la extracción de datos por el *Web Scraping*. Dado que los artículos 71 c) y g) del Decreto Supremo No. 013-2017-JUS advierten que la DGTAIPD tiene la facultad de emitir cualquier lineamiento que sea aplicable en el ámbito de su competencia, consideramos oportuno proveer a los responsables de tratamiento una especie de orientación para la extracción de datos personales mediante el *Web Scraping* conforme con la LPDP (2011) y su Reglamento (2013). Este documento permitiría explicar las implicancias de la evaluación de los principios de tratamiento de la LPDP (2011), además de proporcionar un marco estructurado y claro del cumplimiento de la normativa de protección de datos.

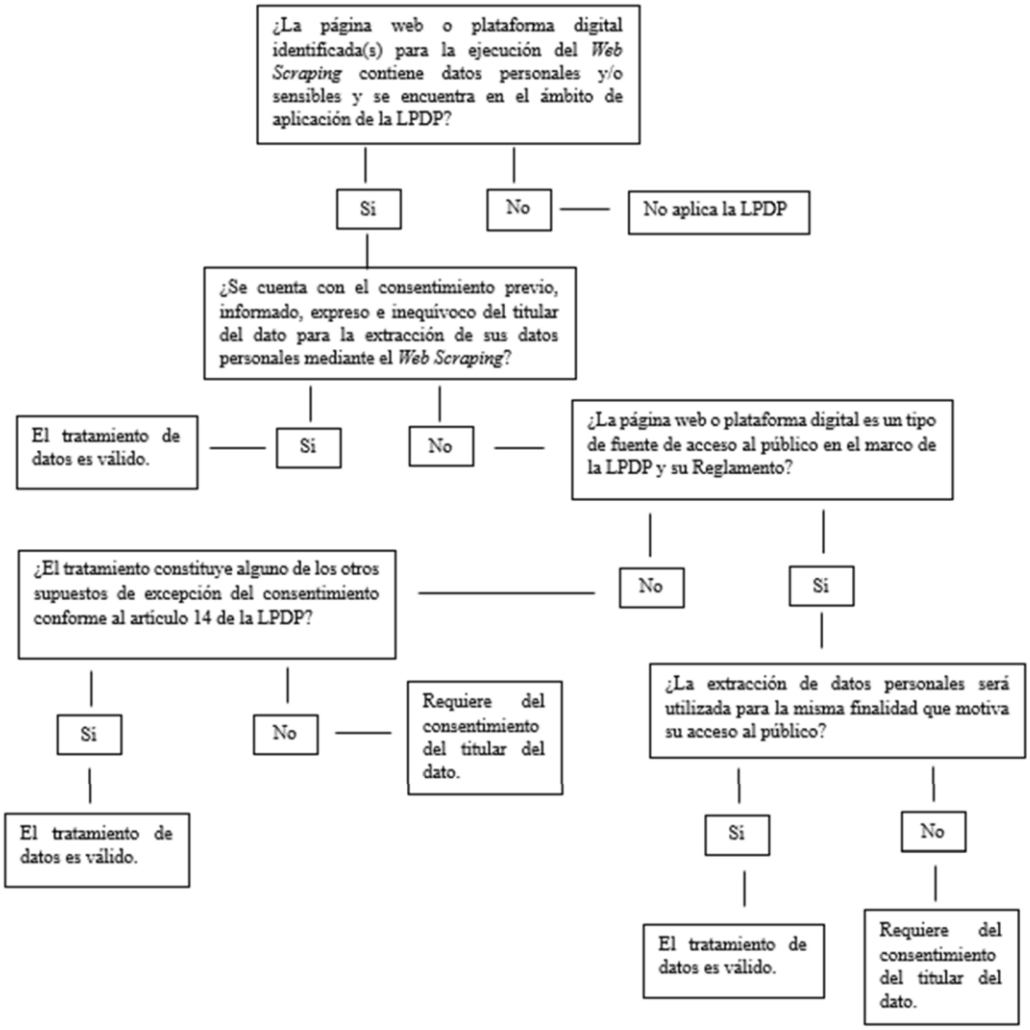
En nuestro caso, el *scraper* tendría una orientación ejemplificada del cómo aplicar los principios tipificados en la normativa, a la vez que una evaluación fáctica evitando

[103] Al respecto, en el artículo 40 del Proyecto de Reglamento de la ANPD se propone la realización de la evaluación de impacto relativo a la protección de datos personales de forma previa al tratamiento, en cuyo caso deberá tomar como referencias a controles y procedimientos técnicos sobre la materia.

[104] Artículo VII. Principios rectores
El titular del banco de datos personales, o en su caso, quien resulte responsable del tratamiento de datos personales, debe cumplir con el régimen jurídico en materia de protección de datos personales de acuerdo con los principios rectores establecidos en el Título I de la Ley; y, asimismo, conforme a los siguientes principios específicos
(...)
2. Principio de responsabilidad proactiva: En el tratamiento de datos personales se deben aplicar las medidas legales, técnicas y organizativas a fin de garantizar el cumplimiento efectivo de la implementación de la normativa de datos personales, y el titular del banco de datos personales o quien resulte responsable, debe ser capaz de demostrar tal cumplimiento.

incumplimientos al amparo de la normativa vigente. Con ánimos de aportar a tal orientación, sin perjuicio de que sea complementado por la DGTAIPD de ser el caso, la evaluación que debería realizar el responsable de tratamiento (*scraper*) seguiría de la siguiente manera:

Figura 21.
Evaluación de tratamiento de datos personales en el marco del *Web Scraping*



Fuente: Elaboración propia

La evaluación presentada constituye un análisis del tratamiento de los datos personales y/o sensibles inmersos cuando ejecutemos una técnica de *Web Scraping*. Inicia con determinar si la página web contiene datos personales y/o sensibles y que, a su vez, nos encontremos dentro de la aplicación de la LPDP (2011) y su Reglamento. De cumplirse ambas condiciones, se prosigue analizando si es que se cuenta con el consentimiento del titular del dato personal para

la extracción de datos personales mediante *Web Scraping*, pues de no ser el caso, seguiríamos preguntándonos si es que la página web o plataforma digital es un tipo de fuente de acceso al público.

Luego, en caso constituya una fuente de acceso al público, entonces cabe preguntarse si es que la finalidad de su disponibilidad es igual a la finalidad por la cual se extraerá la información vía *Web Scraping*. De ser positiva la respuesta, entonces el tratamiento de datos es válido o legal; de lo contrario, requerirá del consentimiento del titular del dato.

La citada evaluación se basa en nuestra legislación peruana de protección de datos personales, así como también en los pronunciamientos de la ANPD cuando se han presentado casos de tratamiento de datos obtenidos por fuentes de acceso público. En contraste del carácter facultativo de la evaluación de impacto relativo al tratamiento de datos personales propuesto en el Proyecto de Reglamento (artículo 40), a nuestro entender consideramos indispensable que, antes de realizar un tratamiento de fuentes de acceso al público, se realice una evaluación de los principios en función de las características de cada caso concreto. Este análisis permitirá cumplir con los principios de la LPDP (2011) y el Reglamento de la Ley 29733 y, con ello, salvaguardar la protección de los titulares frente a tratamientos indebidos de sus datos personales.

De encontrarnos frente al tratamiento lícito de datos personales mediante el *Web Scraping*, entonces corresponderá que se le informe al titular del dato sobre los aspectos de tratamiento conforme con el artículo 18 de la LPDP (2011). Asimismo, un aspecto elemental y primordial para el tratamiento de datos personales mediante la técnica de *Web Scraping* son las medidas de seguridad que se considerarán para salvaguardar tales datos extraídos mediante dicha técnica. Los artículos 9 y 16 de la LPDP (2011) establecen los aspectos sobre la seguridad de la información que deben cumplir los titulares de bancos de datos personales, en cuyo caso deberán adoptar medidas técnicas, organizativas y legales aplicables.

A la fecha, la ANPD cuenta con una Directiva de Seguridad[105] (2013), cuya naturaleza facilitadora y orientativa permite que los responsables de tratamiento puedan conocer y optar por adecuados mecanismos de seguridad acorde al tratamiento de datos vigente. Conforme la

[105] Para mayor información puede acceder al siguiente link: https://cdn.www.gob.pe/uploads/document/file/1401560/Directiva%20de%20seguridad.pdf.

ANPD (Oficio No. 623-2015-JUS/DGPDP), la Directiva de Seguridad es tecnológicamente neutral, por lo que no se puede esperar que establezca "qué hacer" o "cómo hacerlo", pues el resultado, cumplimiento, diseño y las formas de implementación son decisiones que le competen al responsable de tratamiento. Al respecto, el Reglamento, mediante los artículos 39 al 46, realiza un mayor énfasis en materia de seguridad de la información, estableciendo principalmente lo siguiente:

- Control de acceso a la información contenida en los bancos de datos personales: debe identificarse toda la gestión de accesos, de privilegios, identificación del usuario ante el sistema (usuario-contraseña, certificados digitales, tokens) que implica la posibilidad de verificar el tratamiento que realizó cada persona en cada etapa del proceso de tratamiento de datos personales, así como verificaciones periódicas de los privilegios materializado en procedimientos.
- Generar y almacenar registros que contengan la evidencia sobre las interacciones con los datos lógicos: debe permitir la trazabilidad, la información de cuentas de usuario con acceso al sistema, horas de inicio y cierre de sesión. Tales registros deben ser legibles, oportunos y trazables (por ejemplo, el destino de los registros cuando ya no son útiles, su destrucción, transferencia, entre otros).
- Controles de seguridad apropiados: tomando como referencia las recomendaciones de seguridad de la (i) Norma Técnica Peruana NTC-ISO/IEC 27001:2014 Tecnología de la Información. Técnicas de Seguridad. Sistemas de Gestión de Seguridad de la Información. Requisitos. 2a. Edición, desde la vigencia de la Resolución Ministerial No. 004-2016-PCM que aprueba el uso obligatorio de dicha normativa (reemplazo de la NTP ISO/IEC 17799 EDI. Tecnología de la Información) y del (ii) Código de Buenas Prácticas de Seguridad de la Información: ISO 27002.
- Integridad de los datos personales: corresponde la verificación de la integridad de los mismos almacenados en el respaldo, incluyendo la recuperación completa ante una interrupción o daño, garantizando el retorno a su estado anterior.
- La transferencia lógica o electrónica de los datos personales: se realizarán (i) solo por el consentimiento del titular del dato, de ser aplicable, y (ii) tomando en cuenta las medidas necesarias como el cifrado de datos, firmas digitales, información, checksum de verificación, entre otros; a fin de garantizar el no acceso autorizado o pérdida hacia su destino. Ello también aplica para el traslado de datos personales no automatizados.

- Personal autorizado: quienes podrán generar copias o la reproducción de documentos, y acceder a la documentación que aloja los datos personales, a través de mecanismos que permitan identificar los accesos.
- Almacenamiento de documentación no automatizada: los armarios, archivadores u otros objetos que alojen documentos que contengan datos personales deberán contar con todos los mecanismos de seguridad pertinentes (por ejemplo, llave y lugar cerrado cuando no sea necesario el acceso).
- Medidas de seguridad para encargados y responsables de tratamiento: deberán limitar el acceso con su personal, incluyendo el deber de confidencialidad y secreto con motivo al origen de la prestación de los servicios.

De lo señalado en el presente acápite, consideramos fundamental cumplir con una evaluación del tratamiento de datos detallada conforme con los principios de la LPDP (2011) y el Reglamento (2013). Solo así evitaremos transgredir derechos de los titulares de datos, cumpliendo con el principio de información y el de seguridad frente a la técnica del *Web Scraping*.

4.3 Acciones frente a la extracción de datos personales mediante el *Web Scraping*

A pesar de que los datos personales del titular se encuentren disponibles o de fácil acceso (como los perfiles de redes sociales) para la realización de la técnica del *Web Scraping*, corresponderá, en primera instancia, evaluar la posibilidad de su tratamiento a fin de encontrar que no estemos vulnerando alguno de los principios de la LPDP (2011) y Reglamento (2013), en especial el principio de finalidad. De hecho, unos investigadores analizaron tuits públicos para identificar a usuarios de Twitter con problemas de salud mental (Solove, 2006). Sin embargo, los usuarios de Twitter confiaron y se apoyaron en el nivel de privacidad que esperaban les ofreciera dicha red social, pues "*no esperan que su información sea barrida por el scraping de datos*" (Solove, 2006).

Sin perjuicio de que los *scrapers* cumplan con los principios de la LPDP (2011) y su Reglamento (2013) conforme con el acápite anterior, las plataformas digitales y sitios web pueden tomar medidas preventivas de carácter técnico y/o legal que eviten el tratamiento indiscriminado de datos personales mediante el *Web Scraping*.

Una primera medida es la restricción expresa de extracción de datos personales mediante *Web Scraping* (Subirats Maté & Calvo González, 2019) dentro de los Términos y Condiciones. De hecho, tal medida legal se ha implementado en acuerdos institucionales entre las entidades del Estado peruano y organizaciones internacionales, como es el caso del Acuerdo de Cooperación entre el Poder Judicial, Instituto Nacional de Defensa de la Competencia y de la Protección de la Propiedad Intelectual y la Organización Mundial de la Propiedad Intelectual relativo a "WIPO LEX: Sentencias". En tal acuerdo, se señala expresamente que el *Web Scraping* se encuentra rotundamente prohibido bajo cualquier circunstancia (PJ, Indecopi, OMPI, 2020, pág. 3).

Dicha prohibición se utiliza para poder establecer un supuesto de cumplimiento de las finalidades de la página web o plataforma digital entre el propietario del sitio web y el *scraper* (Vanden Broucke y Baesens, 2018). Sin embargo, según Vanden Broucke y Baesens (2018), la simple publicación de tales términos en un sitio web resulta insuficiente para demostrar el cumplimiento de los Términos y Condiciones.

Es así que los "*Browsewrap"* o acuerdos de navegación para tal fin son inadecuados, pues implica la mera publicación del enlace a los Términos y Condiciones para fines de lectura, mas no requiere que los usuarios manifiesten expresamente su consentimiento a los Términos y Condiciones (Dreyer,2013). Ello no permite una aceptación expresa, por parte del *scraper,* para la navegación dentro del sitio web o plataforma digital[106].

La forma correcta de documentar la aceptación de los *scrapers* de la prohibición de ejecutar el *Web Scraping* es utilizando los "*Clickwrap"* o acuerdos de navegación envolvente. Estos métodos requieren que el usuario otorgue su consentimiento a los Términos y Condiciones por medio de un click en un cuadro de diálogo para aceptarlos antes de que el *scraper* pueda continuar utilizando el sitio web[107]. Este debe contar con dos partes: (i) checkbox o casilla de verificación y (ii) un aviso (Dreyer,2013) donde la empresa brinda una indicación de sus acuerdos legales por medio de un texto. Por ejemplo: "*He leído y acepto los términos y condiciones del sitio web".*

[106] Ver Specht, 306 F.3d at 25.
[107] Ver Specht v. Netscape Commc'ns, 306 F.3d 17, 22 n.4 (2d Cir. 2002); Hines v. Overstock.com, 668 F. Supp. 2d 362, 366-67 (E.D.N.Y. 2009).

Como responsable del sitio web y/o plataforma digital, se debe optar por acuerdos de *Clickwrap*, los cuales permiten obtener el consentimiento expreso del *scraper*, siendo que este último deberá cumplir con las disposiciones de tales Términos y Condiciones (Dreyer,2013). En los casos donde no se requiere el registro de una cuenta, se pueden utilizar las casillas de verificación (checkbox) o enlace del tipo "Acepto" donde se incluye de forma expresa la aceptación del *scraper* a los Términos y Condiciones de tal sitio web. En los casos donde se debe crear una cuenta, se incluye un acuerdo explícito de los términos y condiciones de dicho sitio web.

Asimismo, los Términos y Condiciones pueden ser vinculantes cuando el usuario los conoce razonablemente, incluso si el usuario no los conoce antes de que utilice por primera vez el sitio web. Por ejemplo, en caso el usuario acceda numerosas veces al sitio web, los acuerdos del sitio serán aplicables porque el usuario será notificado durante las visitas posteriores a la primera navegación dentro del sitio (Caso Register.com, 356 F.3d, 401-04).

En consecuencia, los responsables de la página web deberán (a) incluir una prohibición expresa de ejecutar el *Web Scraping* dentro de los Términos y Condiciones del sitio web o plataforma digital y (b) el *scraper* deberá respetar dicha restricción los Términos y Condiciones –a. mediante un acuerdo *"Clickwrap"* por medio de un consentimiento expreso o –b. Mediante un acuerdo *"Browsewrap"* si se puede determinar que el usuario los conoce razonablemente por la frecuente utilización del sitio web. Para efectos de una mejor prevención, se sugiere utilizar el *Clickwrap.*

Una segunda medida constituye en incorporar medidas de seguridad dentro de la plataforma digital o página web para proteger los datos personales de sus usuarios. Al respecto, el titular o responsable de tratamiento de la página web o plataforma digital deberá cumplir con las medidas de seguridad necesarias en función a la naturaleza de la información alojada, así como también del contexto, fines de tratamiento, costos incurridos y tecnología disponible.

De acuerdo a la Declaración conjunta sobre *Web Scraping* y protección de datos (Information Commissioner's Office, 2023), se pueden aplicar las siguientes medidas para mitigar los riesgos de extracción de datos vía *Web Scraping:*

- Limitar el número de visitas por hora o día que una cuenta puede realizar a los perfiles de la plataforma y, en caso de detectar una actividad sospechosa, restringir el acceso. Esta medida también permitirá evitar escenarios en donde se sature la plataforma digital perjudicando sus actividades.
- Implementar medidas para detectar *scrapers* revisando patrones de actividad de 'bots'. Sobre ello, los bots maliciosos aprovechan la lógica de la estructura de la página web en lugar de encontrar las vulnerabilidades técnicas, lo que les permite realizar *Web Scraping* de forma desmedida e ilegal (Xiao, 2021).
- Monitorear la rapidez y agresividad que una cuenta busca usuarios nuevos, pues de ser así se deberá restringir el acceso.
- Utilizar métodos para identificar *crawlers*, como CAPTCHAs y bloquear direcciones IP que se han identificado que ejecutan *Web Scraping*.
- Asignar roles específicos al equipo de la organización que permitan identificar todo lo anterior y conozcan los procedimientos de acción y respuesta.
- Notificar cartas de cese o desistimiento en caso se detecte al *scraper* que realiza la actividad de forma ilegal.
- Como parte de una buena práctica, informar a sus usuarios sobre las medidas que han tomado contra el *Web Scraping*, así como mantenerse actualizados con respecto a las nuevas técnicas que permiten *scrapear* datos personales, realizando controles rutinarios y actualizados que mejoren su marco de seguridad de la información.

Con ánimos de cumplir con el principio de seguridad de la LPDP (2011), los propietarios de las páginas web y plataformas digitales (responsables de tratamiento) deberán implementar las medidas legales y técnicas que permitan resguardar los derechos de los titulares de datos personales frente a la técnica del *Web Scraping*, siendo además que podrán fortalecer la confianza de los propios titulares y cumplir con el principio de responsabilidad proactiva tipificada dentro del Proyecto de Reglamento. De igual forma, la implementación de las presentes medidas no exime a que los *scrapers* realicen la evaluación de tratamiento de datos basada en los principios de nuestra LPDP (2011) y el Reglamento (2013), cuyo contenido fue desarrollado y analizado en el acápite anterior.

4.4 Extracción de datos para el entrenamiento de la IA: datos sintéticos

La extracción de datos personales para el entrenamiento de modelos de IA constituye un escenario constante y común en el entorno actual. Solo basta remitirnos a la legislación de Reino Unido, en donde se encuentra en una fase de exploración normativa (consulta pública) con el propósito de encontrar la mejor forma en que los *scrapers* dejen de vulnerar los principios de tratamiento de datos personales cuando realizan extracción masiva de información para el entrenamiento de modelos de IA. Del mismo modo, como se ha mencionado previamente, distintos países han suscrito la Declaración conjunta sobre *Web Scraping* y protección de datos (Information Commissioner's Office, 2023), cuyo contenido establecen medidas unificadoras y preventivas para combatir el tratamiento de datos negligente mediante el *Web Scraping*.

El entrenamiento de modelos de IA es precursor de beneficios para la sociedad que resulta indiscutible de objetar. No obstante, el entrenamiento de los modelos subyacentes al funcionamiento de tales sistemas requiere una enorme cantidad de datos (incluidos datos personales), comúnmente procedentes de una extracción masiva e indiscriminada de datos personales.

En materia de entrenamiento de la IA, los desarrolladores tienen dificultades para ello, destacando la necesidad de conseguir grandes volúmenes de datos para entrenar algoritmos de aprendizaje automático o IA. Muchas veces tales datos no se encuentran disponibles, ya sea porque los datos no existen en cantidades requeridas o porque los datos en cuestión deben representar situaciones que no han ocurrido en la realidad. Asimismo, se requiere clasificar y etiquetar tal cantidad de datos sin utilizar mayores costos o tiempo para conseguir datos anónimos y de calidad. Finalmente, los datos personales presentan restricciones de acceso por motivos de la regulación de protección de datos personales.

Entonces, si los datos sintéticos no cumplen con la utilidad para un fin en específico, no pueden considerarse propiamente como datos sintéticos en el marco de dicho propósito determinado.

Siguiendo con el análisis, para la creación de los datos sintéticos, siempre que en su creación se utilicen datos reales, esta calificaría como una actividad de procesamiento que debe cumplir con lo establecido en la LPDP (2011) y el Reglamento (2013). Por el contrario, si se usan solo datos artificiales, entonces no constituyen datos personales y se excluye la LPDP (2011) y su Reglamento (2013).

El utilizar estos tipos de datos artificiales presentan ciertas ventajas. La primera es que al ser creadas por la tecnología o máquinas el costo de creación y su distribución es bajo, por tanto, facilitan la monetización y la innovación. Adicionalmente, estos datos son utilizados como una forma de anonimización de los datos, en consecuencia, puede aprovecharse para utilizarlos por periodos de uso y conservación más prolongados. De manera adicional, según el informe de datos sintéticos publicado por Alija (2020), presentan resistencia a la reidentificación, puesto que no utilizan datos reales, en consecuencia, no se pueden identificar. Al mismo tiempo, los responsables del tratamiento reducen costos regulatorios, pues no deben preocuparse por la extracción ni tratamiento de estos datos, en tanto son un mecanismo de protección de datos desde el diseño (Hustin, 2019) (Cavoukian, 2010) cuando se trate de casos de uso que necesiten procesar datos personales, porque se añade una capa de protección adicional a los usuarios al no incluir datos personales en su conjunto de información. Por último, agilizan la simulación al generar datos que simulan escenarios que son desconocidos en la vida real o que no han ocurrido (AEPD, 2023).

De manera complementaria, el RGPD (2016) establece que los datos sintéticos no deben contener datos personales, puesto que los datos sintéticos solo conservan propiedades estadísticas o la distribución de los datos personales reales para un fin determinado y, por tanto, evitan el tratamiento de los datos personales (AEPD, 2023). Entonces, ni la generación ni el tratamiento de los datos sintéticos para distintas finalidades vulnera la privacidad de los usuarios (ENISA, 2022). Es decir, se pueden compartir, publicar o tratar de manera más abierta, sin revelar información personal que identifica o hace identificable a un usuario real.

Sin embargo, esta técnica de utilización de datos sintéticos para la extracción de datos para el entrenamiento de la IA también presenta algunas desventajas. En primer lugar, se deben tomar medidas para velar por la calidad y exactitud de los datos. Esto considerando que los datos sintéticos son datos artificiales, que deben analizarse mediante un método de prueba y error para que su uso genere estimaciones más precisas e imparciales con el tiempo. En ese sentido, se debe verificar la calidad de los datos de entrada, pues su calidad y exactitud dependen de estos que pueden originarse de fuentes dispares y del modelo de ajuste de los datos. Aunque sean fáciles de crear, la producción de datos sintéticos debe controlarse, ya que su exactitud no está garantizada. Especialmente en situaciones complejas, la mejor forma de garantizar la calidad de su producción es comparar, a lo largo del tiempo, los resultados de los datos. En

segundo lugar, los datos sintéticos incluyen sesgos tanto en las fuentes de información como en los modelos adoptados (ENISA,2022). Por último, los datos sintéticos no se pueden utilizar tanto para la fase de desarrollo, prueba y validación. Esto porque depender exclusivamente de la IA al generar sus propios datos puede ser perjudicial, pues en caso haya errores, estos se multiplicarán una vez se generen con los datos artificiales erróneos nuevos datos artificiales los cuales tendrán las mismas incongruencias.

En suma, a la fecha, la generación de datos sintéticos no es una opción definitiva para el tratamiento masivo de datos personales mediante *Web Scraping*, en tanto se debe evaluar su utilización y oportunidad caso por caso. Sin embargo, en un futuro cercano, se pronostica que las empresas tecnológicas agoten la información en la nube o Internet para el desarrollo de modelos de IA; lo que los obligará a crear sus propios datos artificiales para continuar desarrollando nuevos productos o servicios y a entrenar a sus desarrollos de IA con datos como textos y gráficos de otros modelos de IA.

CONCLUSIONES

En base a lo analizado en el presente trabajo, llegamos a las siguientes conclusiones:

1. Desde los inicios del Internet, se tuvo la propuesta que sea una infraestructura de comunicaciones que permita el acceso y alcance de información interconectada a un solo *click*. Eventualmente, tal propuesta evolucionó y las necesidades de desarrollar nuevos actores para el sostenimiento del Internet han permitido que sea considerado como principal partícipe en el despliegue de la Cuarta Revolución Industrial, cuyo fenómeno, a su vez, ha permitido la creación de nuevas tecnologías emergentes y, con ello, la valorización exponencial de los datos personales. Es así que los datos cumplen un rol fundamental para el desarrollo de nuevas tecnologías, pues constituyen el oxígeno de la subsistencia y desarrollo de las páginas web y/o plataformas digitales.
2. Dentro de esta esfera de constante evolución y emisión de datos personales, existen técnicas que permiten extraer los mismos de forma automatizada, entre los cuales se encuentra el *Web Scraping.* El *Web Scraping* es una técnica automatizada de extracción de datos personales dentro de páginas web y/o plataformas digitales. Para ello, el flujo del mismo inicia con el análisis de la página web para luego pasar al *crawling* (o también llamado rastreo), y con ello, al *Web Scraping*, es decir, a la extracción de los datos personales. Finalmente, el proceso termina con la organización y/o recopilación de datos personales en función a la finalidad que ha motivado el *Web Scraping*. La presente técnica es comúnmente usada vía Python, como lenguaje de programación, en conjunto con Jupyter Notebook y la librería Selenium. Se ha demostrado, vía simulación, que una página web con acceso a datos disponibles, como la página web de Proyectos de Ley del Congreso, es susceptible de ejecutar la técnica de *Web Scraping*. Del mismo modo, mediante la IA es factible construir códigos funcionales al *Web Scraping*, considerando además que existe una tendencia a utilizar herramientas IA para la construcción de soluciones y productos finales.
3. El derecho a la autodeterminación informativa consiste en el control que toda persona goza sobre sus datos personales, data relacionada a la intimidad personal y familiar, así como toda aquella que sea de control y reserva. A pesar de la deficiencia normativa por su reconocimiento constitucional, nuestro Tribunal Constitucional, a través de distintas sentencias, ha advertido que el presente derecho se encuentra tipificado en el artículo 2.6 de nuestra Constitución.

4. De acuerdo a documentos y tratados internacionales, los datos personales constituyen toda información de una persona identificada o identificable. Para nuestra LPDP (2011) y Reglamento, los datos personales son aquellos datos que identifican o son identificables a una persona natural vía medios razonablemente utilizados. A su vez, nuestra legislación realiza una clasificación y determina a los datos sensibles aquella que abarca (i) los datos biométricos y (ii) aquellos que por sí solos hacen posible la identificación del titular, es decir, no se requiere de ninguna otra herramienta para la identificación. Estos datos cuentan con una protección mayor por la calidad de su contenido.
5. El responsable del tratamiento es quien decide el tratamiento de los datos personales, aunque no estén alojados en un banco de datos personales. Asimismo, el tratamiento de datos personales constituye cualquier procedimiento técnico, automatizado o no, que permite, entre otros tipos, la extracción de datos personales. Cuando se realiza la técnica del *Web Scraping*, se está realizando un tratamiento de datos personales en sí, en cuyo caso deberá cumplirse con los principios y obligaciones tipificadas en la LPDP (2011) y su Reglamento.
6. En el marco internacional, el *Web Scraping* ha sido detectado por distintas jurisdicciones como técnica vulneradora a los principios de privacidad. Tan es así que existen casos emblemáticos en España, Italia, Reino Unido y Colombia, en donde la Autoridad de Protección de Datos Personales (o su equivalente) ha terminado sancionando al *scraper* por no haber obtenido el debido consentimiento del titular y, en consecuencia, transgrediendo la normativa de privacidad aplicable. Es por ello que se ha logrado firmar una Declaración conjunta sobre Web Scraping y protección de datos (Information Commissioner's Office., 2023) con distintos países de la región, cuyo contenido fija una serie de recomendaciones frente al tratamiento indiscriminado de datos personales, así como el papel que cumplen los propietarios de las páginas web y/o plataformas digitales frente al *Web Scraping*. Al respecto, se reconoce el grado de responsabilidad de tales propietarios para salvaguardar los datos personales alojados en sus plataformas, quienes deberán implementar medidas técnicas y/o legales que eviten el incumplimiento de la normativa de datos personales correspondiente.
7. En el marco nacional, reconocemos el gran impacto que contiene el *Web Scraping* en la privacidad, pues dada la constante extracción de datos personales de distintas páginas y/o plataformas digitales de forma constante y automatizada, este tratamiento puede

vulnerar los principios básicos de la LPDP (2011) y su Reglamento (2013), en especial el principio de finalidad y del consentimiento.

8. La solución ante dicho problema, y que nos motivó a redactar el presente trabajo, no consiste en regular, *per se*, el *Web Scraping*, pues dicha técnica se subsume a uno de los tipos de tratamiento de datos personales tipificados en la LPDP (2011) y su Reglamento (2013). Además, establecer una medida que regule en sí dicha técnica ralentizará la innovación y, del mismo modo, generaría un escenario de insuficiencia regulatoria al existir normativa aplicable a su naturaleza.
9. Usualmente se extraen datos personales mediante la técnica del *Web Scraping* desde fuentes de acceso al público o fuentes disponibles en las páginas web y/o plataformas. Así, el artículo 14 de la LPDP (2011) enlista los supuestos en los que no se requerirá del consentimiento del titular para la ejecución del tratamiento de datos. Entre esos supuestos está el tratamiento de datos obtenidos por fuentes de acceso al público (inciso 2, artículo 14).
10. En aras de proteger los datos personales de los titulares frente a la indiscriminada extracción de datos personales mediante la técnica del *Web Scraping*, proponemos lo siguiente:

    A) Una modificación parcial del artículo 14 inciso 2 de la LPDP (2011), cuyo contenido resalta la aplicación de los principios de la LPDP (2011) y su Reglamento (2013), en supuestos donde (i) nos encontremos frente a una fuente de acceso al público y (ii) en función a las circunstancias y caso concreto. Esta evaluación brindará mayor seguridad jurídica al reconocerse expresamente en la LPDP (2011) y, a su vez, será realizado *ex ante* del tratamiento de datos personales. Aunado a ello, se requerirá documentar dicho tratamiento para que la ANPD pueda, de oficio o a pedido de parte, revisar el análisis del *scraper* que lo motivó a ejecutar el *Web Scraping*. Documentar este procedimiento se alinea con el principio de responsabilidad proactiva soslayada en la Propuesta de Reglamento, y con el principio de seguridad desarrollado con mayor énfasis en la Directiva de Seguridad.

    B) De manera complementaria y con el propósito de orientar de forma práctica el análisis para la aplicación de los principios de la LPDP (2011), proponemos una evaluación de tratamiento de datos en el marco del *Web Scraping*. Dicho esquema permitirá conocer al *scraper* los pasos a seguir en materia de datos

personales, así como también incentivar el cumplimiento de la normativa correspondiente. Conforme a los pronunciamientos de la ANPD, no podrá realizarse tratamiento de datos obtenidos de fuentes de acceso al público para una finalidad distinta a la que motivó su disponibilidad o publicación, pues de lo contrario, amerita obtener el consentimiento previo, expreso, informado e inequívoco del titular del dato.

11. De esta manera, si bien el *scraper* se encuentra frente a una excepción del consentimiento, deberá cumplir con los otros principios tipificados en la LPDP (2011), tales como el deber de información, en donde se le informará al titular de datos personales sobre los alcances del tratamiento, de acuerdo con el artículo 18 de la LPDP (2011). Asimismo, deberá cumplir con el principio de seguridad, cumpliendo con las disposiciones de la LPDP (2011) y orientándose bajo la Directiva de Seguridad.
12. Con la finalidad de proteger a los titulares de datos personales frente a la ejecución negligente del *Web Scraping* y considerando que los propietarios de las páginas web y/o plataformas digitales son los responsables de los datos personales alojados en su plataforma, dichos propietarios deberán implementar las medidas técnicas (p.e. implementar CAPTCHAS o incorporar bots que detecten actividades de *Web Scraping*) y legales (p.e. prohibir la práctica del *Web Scraping* en sus Términos y Condiciones) para salvaguardar los derechos de los titulares de datos. De esta manera, podrán cumplir con el principio de seguridad de la LPDP (2011) y responsabilidad proactiva tipificado dentro del Proyecto de Reglamento.
13. Una posible propuesta que evita la extracción de datos personales obtenidos por fuentes de acceso al público vía *Web Scraping* para el entrenamiento de soluciones de IA es la utilización de datos sintéticos. Este tipo de datos son creados artificialmente para usos específicos y, por ello, son irreales. Sin embargo, se deben utilizar para casos concretos en específico teniendo en cuenta los problemas que puede generar por los errores que la IA no controla completamente. En ese sentido, se estima que en un futuro se llegue a utilizar con plenitud con la finalidad de seguir entrenando productos de la IA.

REFERENCIAS BIBLIOGRÁFICAS

Abad, S. (1994). Hábeas data y conflicto entre órganos constitucionales. En Comisión Andina de Juristas, Lecturas sobre Temas Constitucionales 10. Lima.

Acquisti, A., Taylor, C., & Wagman, L. (2016). The economics of privacy. In *Journal of Economic Literature* (Vol. 54, Issue 2, pp. 442–492). American Economic Association. https://doi.org/10.1257/jel.54.2.442

ACLU v. Clearview AI, Inc. (2020). No. 2020 CH 04353.

Acuerdo Nacional (2017). Sociedad de la información y sociedad del conocimiento. https://acuerdonacional.pe/politicas-de-estado-del-acuerdo-nacional/politicas-de-estado/politicas-de-estado-castellano/iv-estado-eficiente-transparente-y-descentralizado/35-sociedad-de-la-informacion-y-sociedad-del-conocimiento/.

Acuerdo de Cooperación entre la Organización Mundial de la Propiedad Intelectual, el Poder Judicial del Perú y el Instituto Nacional de Defensa de la Competencia y de la Protección de la Propiedad Intelectual relativoo a "Wipo Lex: Sentencias", (2020). https://www.pj.gob.pe/wps/wcm/connect/7c8f9b00452338c2963ade807c1f73f9/ACUERDO+INDECOPI-PODER+JUDICIAL-OMPI+BASE+DE+DATOS+WIPO-LEX+%E2%80%93+SUSCRITO+%281%29.pdf?MOD=AJPERES&CACHEID=7c8f9b00452338c2963ade807c1f73f9

Adinolfi, G. (2006). Autodeterminación Informativa, El europeísmo español vs El nacionalismo italiano: consideraciones acerca de un principio general y derecho fundamental. Revista Española de Derecho Constitucional. https://dialnet.unirioja.es/descarga/articulo/2233714.pdf

Adler, M.J. (1986). A Guidebook to Learning: For a Lifelong Pursuit of Wisdom. Macmillan, London.

Agencia de la Unión Europea para la Ciberseguridad (ENISA). (2022). Ingeniería de la protección de datos (P. D. y M. A. (Agencia de la U. E. para la Ciberseguridad), Eds.). https://doi.org/10.2824/09079

Agencia Española de Protección de Datos. (2008). Informe 0342/2008. https://www.aepd.es/documento/2008-0342.pdf

Agencia Española de Protección de Datos. (2018). Guía para el cumplimiento del deber de

informar. https://www.aepd.es/guias/guia-modelo-clausula-informativa.pdf

Agencia Española de Protección de Datos. (2019). Guía de Privacidad desde el Diseño. https://www.aepd.es/sites/default/files/2019-11/guia-privacidad-desde-diseno.pdf

Agencia Española de Protección de Datos. (2019). Resolución de Procedimiento Sancionador del Procedimiento No. PS/00240/2019. https://www.aepd.es/documento/ps-00240-2019.pdf

Agencia Española de Protección de Datos. (2023). *Datos sintéticos y protección de datos*. https://www.aepd.es/prensa-y-comunicacion/blog/datos-sinteticos-y-proteccion-de-datos

Aguilar, H. (2023). Raspando la Arqueología: Una Aproximación Metodológica desde el Web Scraping y Text Mining. Revista Del Museo de Antropología, 16(3), 439–450. https://doi.org/10.31048/1852.4826.v16.n2.41094

Alarcón-Urbistondo, P., Rojas-de-Gracia, M.-M., & Casado-Molina, A. (2023). Proposal for Employing User-Generated Content as a Data Source for Measuring Tourism Destination Image. *Journal of Hospitality & Tourism Research*, 47(4), 643-664. https://doi.org/10.1177/10963480211012756

Alexy, R. (2002). Teoría de los derechos fundamentales. Centro de Estudios Constitucionales. Madrid, 86. https://www.pensamientopenal.com.ar/system/files/2014/12/doctrina37294.pdf

Alija, A. (2023). *Datos Sintéticos: ¿Qué Son Y Para Qué Se Usan?* Secretaria de Estado de Digitalización e Inteligencia Artificial. https://datos.gob.es/sites/default/files/doc/file/informe-datos-sinteticos-es.pdf

Álvarez, S. (2021). La protección de la vida privada y familiar: Sexualidad, reproducción y violencia. España: MARCIAL PONS, EDICIONES JURÍDICAS Y SOCIALES.

Anzures Gurría, J. J. (2020). The internet as a human right. Legal nature an purposes. *Boletin Mexicano de Derecho Comparado*, *53*(158), 521–552. https://doi.org/10.22201/iij.24484873e.2020.158.15628

Anureev, A. A. (2015). Operational semantics development for procedural programming languages based on conceptual transition systems. *NCC Bulletin*, Volumen 38, 1-28. https://nccbulletin.ru/files/article/anureev_bulleten_2015.pdf

ANPD. Oficio No. 140-2014-JUS/DGPDP de fecha 21 de marzo de 2014. https://cdn.www.gob.pe/uploads/document/file/1463947/Representante%20de%20persona%20jur%C3%ADdica.pdf

ANPD. Oficio No. 623-2015-JUS/DGPDP de fecha 23 de diciembre de 2015. https://cdn.www.gob.pe/uploads/document/file/1466174/Alcances%20sobre%20algunas%20de%20las%20orientaciones%20contenidas%20en%20la%20Directiva%20de%20Seguridad..pdf

ANPD. (2014). El Derecho Fundamental a la Protección de Datos Personales – Guía para el Ciudadano. Lima, Perú. https://cdn.www.gob.pe/uploads/document/file/1401558/El%20derecho%20fundamental%20a%20la%20protecci%C3%B3n%20de%20datos%20personales.pdf

ANPD. Oficio No. 635-2015-JUS/DGPDP de fecha 22 de diciembre de 2015. https://cdn.www.gob.pe/uploads/document/file/1466173/Diferencias%20entre%20las%20terminolog%C3%ADas%20%C2%ABbancos%20de%20datos%20personales%C2%BB%20y%20bases%20de%20datos..pdf

ANPD. Oficio No. 213-2017-JUS/DGTAIPD de fecha 10 de octubre de 2017. https://cdn.www.gob.pe/uploads/document/file/1465782/Tratamiento%20de%20datos%20personales%20contenidos%20en%20fotograf%C3%ADas..pdf

ANPD. Opinión Consultiva No. 034-2021-JUS/DGTAIPD de fecha 07 de septiembre de 2021. https://cdn.www.gob.pe/uploads/document/file/2192913/Sobre%20las%20obligaciones%20del%20encargado%20y%20el%20responsable%20del%20tratamiento%20de%20datos%20personales.pdf?v=1632237699

ANPD. (2019). Guía práctica para la observancia del deber de informar. Aprobada mediante Audiencia Nacional. (2012). Sentencia de 31 de mayo de 2012, núm. JUR/2012/221309. Ff. jj, CUARTO.

Resolución Directoral No. 80-2019-JUS/DGTAIPD. https://www.gob.pe/institucion/minjus/informes-publicaciones/353793-guia-practica-para-la-observancia-del-deber-de-informar

Asamblea General de las Naciones Unidas. (1948). Declaración Universal de los Derechos Humanos. https://www.un.org/es/universal-declaration-human-rights/.

Asia-Pacific Economic Cooperation (APEC). (2005). *Marco de Privacidad APEC.*

https://www.apec.org/docs/default-source/publications/2005/12/apec-privacy-framework/05_ecsg_privacyframewk.pdf?sfvrsn=d3de361d_1

Balkin, J. (2016). Information Fiduciaries and the First Amendment.

Banco Interamericano de Desarrollo (BID). (s.f.). *Manual de orientación para participar en redes sociales* (p. 11). https://publications.iadb.org/es/publicacion/14832/manual-de-orientacion-para-participar-en-redes-sociales

Bedi, M. (2017). The Fourth Amendment disclosure doctrines. *William & Mary Bill of Rights Journal, 26*(2), 461-482.

Baskaran, U., & Ramanujam, K. (2018). Automated scraping of structured data records from health discussion forums using semantic analysis. *Informatics in Medicine Unlocked, 10*, 149–158. https://doi.org/10.1016/j.imu.2018.01.003

Belyakova, J. (2016). Language Support for Generic Programming in Object-Oriented Languages: Design Challenges. *Trudy ISP RAN / Proc. ISP RAS, 28*(2), 5-32. https://julbinb.github.io/files/papers/ispras2016_OO-generics.pdf

Proyecto de Ley 220. (2018). Ohio Data Protection Act.

Blanco Garrido, C. (2018). Extracción De Características Para Algoritmos De Aprendizaje Automático Aplicado Al Reconocimiento De Vehículos. [Escuela Técnica Superior de Ingenieros Industriales (UPM)]. http://oa.upm.es/50271/1/TFG_CRISTINA_BLANCO_GARRIDO.pdf

Bosch, O. ten, Windmeijer, D., Van Delden, A., & Van den Heuvel, G. (2018). Web scraping meets survey design: Combining forces. Bigsurv18 Conference, October 1–13.

Bundesdatenschutzgesetz für den Bereich des Gesundheitswesens, 1995, BGBl. I S. 2950.

Cámara de Diputados del H. Congreso de la Unión. (2010). *Ley Federal de Protección de Datos Personales en Posesión de los Particulares*. Secretaría General, Secretaría de Servicios Parlamentarios. Recuperado de https://www.diputados.gob.mx/LeyesBiblio/pdf/LFPDPPP.pdf

Cámara de Diputados del H. Congreso de la Unión. (2011). *Reglamento de la Ley Federal de Protección de Datos Personales en Posesión de los Particulares*. Secretaría General, Secretaría de Servicios Parlamentarios. Recuperado de https://www.diputados.gob.mx/LeyesBiblio/regley/Reg_LFPDPPP.pdf

Camargo-Henríquez, I.; Núñez-Bernal, Y. (2022). A Web Scraping based approach for data research through social media: An Instagram case. *V Congreso Internacional en Inteligencia Ambiental, Ingeniería de Software y Salud Electrónica y Móvil (AmITIC)*, 1-4.

Cambridge University Press. (s.f.). Data. In Cambridge English Dictionary. https://dictionary.cambridge.org/dictionary/english/data

Calhoun v. Google LLC. (2021). Case No. 20-CV-05146-LHK. https://caselaw.findlaw.com/court/us-dis-crt-n-d-cal-san-jos-div/2118953.html

California Office of the Attorney General. (s.f.). California Consumer Privacy Act (CCPA) Fact Sheet. Recuperado de https://oag.ca.gov/sites/all/files/agweb/pdfs/privacy/ccpa-fsor.pdf

Castillo Jiménez, C. (1994). Estatuto de la agencia de protección de datos, Informática y derecho: Revista iberoamericana de derecho informático, nº 6-7.

Cavoukian, A. (2010). Identity in the Information Society (2010). *Privacy by design: The definitive workshop. A foreword.* Volumen 3, Número 2, pp 247-251 https://link.springer.com/content/pdf/10.1007/s12394-010-0062-y.pdf

Cavero, E., & Holguín, C. (2013, June). Protección de datos personales: naturaleza, principios rectores y derechos subjetivos. El derecho de autodeterminación informativa en la ley peruana. *Revista Actualidad Jurídica*, (235).

Centro de Ciberseguridad del Gobierno de Buenos Aires. (2024). Web scraping: La maravilla de los datos en la web. https://buenosaires.gob.ar/noticias/web-scraping-la-maravilla-de-los-datos-en-la-web

Chan Sánchez, J. (2008). APEC y el Perú: guía sobre el foro de Cooperación Económica Asia-Pacífico y la participación del Perú. Perú: Universidad de Lima, Fondo Editorial.

Chanamé Orbe, R. (2003). Hábeas data y el derecho fundamental a la intimidad de la persona. Lima: UNMSM. Biblioteca de la Facultad de Derecho y Ciencia Política.

Chapman and Hall/CRC. (2012). *How Does the Internet Work?* (2nd ed.). https://doi.org/10.1201/b12327-15

Chiriboga, G. (2001). La acción de amparo y de hábeas data: garantías de los derechos constitucionales y su nueva realidad jurídica. Quito, AAJ/ILDIS.

Christensen, C.M. (1997) The Innovator's Dilemma. Harvard Business Review Press, Cambridge, MA.

Código Civil del Perú. (1984).

Código Civil de California (s.f.)

Colorado Revised Statutes. (2021). § 6-1-1303(24).

Comisión Andina de Juristas. (1997). Protección de los derechos humanos: definiciones operativas (p. 182). Lima.

Comisión de Ciencia Innovación y Tecnología (2023). Dictamen del Proyecto de Ley 2775/2022-CR, Ley que promueve el uso de la Inteligencia Artificial en favor del desarrollo económico y social del país. Congreso de la República. https://wb2server.congreso.gob.pe/spley-portal-service/archivo/ODQ1OTA=/pdf

Consejo de Europa. (1981). Convenio para la protección de las personas con respecto al tratamiento automatizado de datos de carácter personal.

Constitución de la Nación Argentina (1994). 1° ed. Editorial legislativa.

Congreso de la República de Colombia. (1991). Constitución Política de Colombia de 1991.

Constitución Española. (1978). Boletín Oficial del Estado.

Constitución Política del Perú, (1993).

Congreso Constituyente Democrático. (1998). *Diario de los Debates: Debate Constitucional Pleno 1993* (publicación oficial); julio 1998; tomo I; Págs. 111-112.Congreso de la República del Perú. (s.f.). https://wb2server.congreso.gob.pe/spley-portal/#/expediente/search

Congreso de la Nación Argentina. (2014). *Código Civil y Comercial de la Nación, Aprobación*. Fecha de sanción 01-10-2014. Publicada en el Boletín Nacional del 08-Oct-2014. https://www.argentina.gob.ar/normativa/nacional/ley-26994-235975

Connecticut Public Acts. (2023). No. 22-15, § 1(27).

Corte Constitucional de Colombia. (1992). Sentencia No. T-414/92.

Corte Constitucional de Colombia (2022). Sentencia T-729.

Corte Constitucional de Colombia. (2011). Sentencia C-748.

Cox Broadcasting Corp. v. Cohn. (1975). 420 U.S. 469. https://supreme.justia.com/cases/federal/us/420/469/

Davara, Miguel (1993). Derecho Informático, Editorial Aranzadi - Pamplona, Edición España, Pág. 50.

Davara Fernández de Marcos, I. (2011). Hacia la estandarización de la protección de datos personales: propuesta sobre una "tercera vía o tertium genus" internacional. Las Rozas: La Ley.

Delgado, (s.f). Cómo será el Futuro de la Inteligencia Artificial: Mas allá de la ficción. https://books.google.com.pe/books?id=HVfqEAAAQBAJ&printsec=frontcover&hl=es#v=onepage&q&f=false

Dreyer, B. Y. A. J. (2013). Internet '*Data Scraping*'. New York Law Journal, ALM Media Properties, LLC. https://www.skadden.com/-/media/files/publications/2014/01/070071319-skadden.pdf

Decreto de Urgencia No. 006-2020. Decreto de Urgencia que crea el Sistema Nacional de Transformación Digital. Diario Oficial El Peruano (2020). https://busquedas.elperuano.pe/dispositivo/NL/1844001-1

Decreto de Urgencia No. 007-2020. Por medio del cual se aprueba el marco de confianza digital y dispone medidas para su fortalecimiento. Diario Oficial El Peruano (2020). https://cdn.www.gob.pe/uploads/document/file/2790485/Decreto%20de%20Urgencia%20N%C2%BA%20007-2020.pdf?v=1643322610

Decreto Legislativo No. 1412. Decreto Legislativo que aprueba la Ley de Gobierno Digital. Diario Oficial El Peruano (2018). https://www.gob.pe/institucion/pcm/normas-legales/289706-1412

Decreto No. 1558/2001. Decreto que Reglamenta la Ley 25326, Ley de Protección de datos personales. (2001). https://www.argentina.gob.ar/normativa/nacional/decreto-1558-2001-70368

Decreto Reglamentario parcial No. 1377. Por el cual se reglamenta parcialmente la Ley 1581 de 2012, Derogado Parcialmente por el Decreto 1081 de 2015. (2013). https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=53646

Decreto Legislativo 30 Giugno (2003), n 196. recante il "Codice in materia di protezione dei dati personali". https://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:decreto.legislativo:2003-

06-30;196.

Decreto Supremo No. 003-2013-JUS. Aprueban Reglamento de la Ley No. 29733, Ley de Protección de Datos Personales. Diario Oficial El Peruano (2013). https://spij.minjus.gob.pe/spij-ext-web/#/detallenorma/H1075450

Decreto Supremo No. 013-2017-JUS. Aprueba el Reglamento de Organización y Funciones del Ministerio de Justicia y Derechos Humanos. Diario Oficial El Peruano (2017). https://www.gob.pe/institucion/minjus/normas-legales/1439578-013-2017-jus

Decreto 1074, Decreto Único Reglamentario del Sector Comercio, Industria y Turismo (2015)

Defensoría del Pueblo. (2019). Manual de Protección de Datos Personales. https://www.defensoria.gob.pe/wp-content/uploads/2019/11/Manual-de-Protecci%C3%B3n-de-Datos-Personales.pdf

Desantes Guanter, J. (1990). Los límites de la información. La información en la jurisprudencia del Tribunal Constitucional: las 100 primeras sentencias.

DGTAIPD. (2018). Resolución Directoral No. 453-2018-JUS/DGTAIPD de fecha 12 de marzo de 2018. https://historial.pe/pdf/EL_COMERCIO_DESINDEXACION.pdf

DGTAIPD. (2019). Resolución Directoral No. 84-2019-JUS/DGTAIPD de fecha 03 de diciembre de 2019. https://cdn.www.gob.pe/uploads/document/file/2011717/RD%2084-2019-DGTAIPD.pdf.pdf

DHI Grp., Inc. v. Kent, No. (2017). CV H-16-1670, 2017 WL 4837730, at *2–4.

Diario Oficial de la Unión Europea. (1995). Directiva 95/46/CE del Parlamento Europeo y del Consejo, de 24 de octubre de 1995, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos. https://www.boe.es/buscar/doc.php?id=DOUE-L-1995-81678#:~:text=Los%20Estados%20miembros%20prohibir%C3%A1n%20el,la%20salud%20o%20a%20la%20sexualidad3.5

Diario Oficial de la Unión Europea (2009). Tratado de Lisboa por el que se modifican el Tratado de la Unión Europea y el Tratado Constituido de la Comunidad Europea. https://www.boe.es/doue/2007/306/Z00001-00271.pdf

Diario Oficial de la Unión Europea (2016). Reglamento UE 2016/679, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE [RGPD]. https://www.boe.es/doue/2016/119/L00001-00088.pdf

Dictamen Nº 15/04, Proyecto de “Reglamentación general del tratamiento de la información judicial”, Expte. 129/2002. (2004, diciembre 9). Dictamen DNPDP N° 69/04. Buenos Aires.

https://www.argentina.gob.ar/sites/default/files/d2004_069.pdf

Diouf, R., Sarr, E. N., Sall, O., Birregah, B., Bousso, M., & Mbaye, S. N. (2019). *Web scraping: State-of-the-art and areas of application*. In *Proceedings of the 2019 IEEE International Conference on Big Data* (pp. 6040-6042). IEEE. https://doi.org/10.1109/BigData47090.2019.9005594

Duran, Ana. (2015). La figura del responsable en el derecho a la protección de datos. Génesis y evolución normativa ante el cambio tecnológico y en perspectiva multinivel. [Tesis doctoral]. Universitat Autònoma de Barcelona. https://ddd.uab.cat/pub/tesis/2015/hdl_10803_319454/abdc1de1.pdf

El, E. N., Circuito, T. D. E., & Condado, D. E. L. (2020). En el tribunal de circuito del condado de Cook, Illinois.

Eguiguren Praeli, F. J. (2002). Estudios Constitucionales. ARA editores, 1era edición.

Eguiguren Praeli, F. J. (2004). La Libertad de Expresión e Información y el Derecho a la Intimidad Personal. Su desarrollo actual y conflictos, Palestra.

Eguiguren Praeli, F. (2004). Libertades de expresión e información, intimidad personal y autodeterminación informativa: contenido, alcances y conflictos (Tesis de Magíster, Pontificia Universidad Católica del Perú). Repositorio Institucional PUCP.

Eguiguren Praeli, F. J. (2015). El derecho a la protección de los datos personales. Algunos temas relevantes de su regulación en el Perú. *THEMIS Revista De Derecho*, (67), 131-140. https://revistas.pucp.edu.pe/index.php/themis/article/view/14462

EUR-Lex. (2023). Reglamento (UE) 2023/2854 del Parlamento Europeo y del Consejo de 14 de noviembre de 2023 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por las instituciones, órganos y organismos de la Unión y a la libre circulación de estos datos, y por el que se deroga el Reglamento (UE) 2018/1725. https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A32023R2854

Estados Unidos. (2000). Principios de Puerto Seguro.

Red Iberoamericana de Protección de Datos (2017). Estándares de protección de datos personales para los Estados Iberoamericanos.

Feist Publications, Inc. v. Rural Tel. Serv. Co. (1991). 499 U.S. 340. de: https://supreme.justia.com/cases/federal/us/499/340/

Fernández Naranjo, M. (2023). Benitez López, A., Inteligencia Artificial en perspectiva.

Contrastes. Revista Internacional De Filosofía, *28*(3), 161–163. https://doi.org/10.24310/contrastes.28.3.2023.16426

Fernández Segado, F. (1997). El Régimen Jurídico del Tratamiento automatizado de los datos de carácter personal en España. Ius et Praxis.

Fernández-Aller, C., & Serrano Pérez, M. M. (2022). ¿Es posible una Inteligencia artificial respetuosa con la protección de datos? Doxa. Cuadernos de Filosofía Del Derecho, 45, 307. https://doi.org/10.14198/doxa2022.45.11

Ferrara, E., De Meo, P., Fiumara, G., & Baumgartner, R. (2014). Web data extraction, applications and techniques: A survey. *Knowledge-Based Systems*, *70*, 301-323. https://doi.org/10.1016/j.knosys.2014.07.007

Ferrero, G. (2023). Responsabilidad de los Internet Service Providers (ISP) por violación al derecho de autor. En Tribunal de Justicia de la Comunidad Andina (TJCA) y la Asociación Interamericana de la Propiedad Intelectual (ASIPI) (Eds.), La propiedad industrial y el derecho de autor en Iberoamérica (pp. 123-145). Editorial Andina. https://ejemplo.com/libro-digital

Flores, P. (1987). Diccionario de términos jurídicos. V. 3. Lima: Marsol Editores.

Flores, R., & Dapkevicius, Ruben. (2004). Amparo, Hábeas Corpus y Hábeas Data. B de F. Montevideo – Buenos Aires, Argentina.

Florida Star v. B.J.F., 491 U.S. 524 (1989). In Florida Star, the plaintiff rape victim asserted a claim under a Florida law banning publication of the name of a sexual offense victim. The Court struck down this law on First Amendment grounds. Florida Star acknowledged that privacy interests may sufficiently override First Amendment interests, but, in the instant case, found the means employed were not narrowly tailored to furthering the privacy interests. Id. at 537. https://supreme.justia.com/cases/federal/us/491/524/

Floridi, L. (2008). 'Data', in W.A. Darity (ed.), International Encyclopedia of the Social Sciences, 2nd edition. Detroit: Macmillan.

Floridi, L. (2010). Information: A Very Short Guide. Oxford University Press, Oxford.

Frigerio Dattwyler, C. (2018). Mecanismos de regulación de datos personales: una mirada desde el análisis económico del derecho. Revista chilena de derecho y tecnología, 7(2), 45-80. https://dx.doi.org/10.5354/0719-2584.2018.50578

Fundación Telefónica (2023). Telos 123. Inteligencia artificial: Un punto de inflexión en la humanidad. https://www.fundaciontelefonica.com.pe/cultura-digital/publicaciones/telos-123-

inteligencia-artificial/804/

Gacitúa Espósito, A. L. (2014). El Derecho fundamental a la protección de datos personales en el ámbito de la prevención y represión penal europea: (en busca del equilibrio entre la libertad y la seguridad). España: Universitat Autònoma de Barcelona.

Gaspar, W. B., Magrani, B., Ferraz, J. V., & Souza, C. A. P. de. (2012). Filtrado de contenido en América Latina: razones e impacto en la libertad de expresión. Hacia una internet libre de censura: propuestas para América Latina. http://www.palermo.edu/cele/pdf/internet_libre_de_censura_libro.pdf%0Ahttp://www.palermo.edu/cele/libertad-de-expresion/publicaciones.html

García Belaunde, D. (1994). Garantías constitucionales en la Constitución Peruana de 1993. En Comisión Andina de Juristas, La Constitución de 1993; análisis y comentarios (Serie Lecturas sobre Temas Constitucionales No. 10, págs. 259-260). Lima.

García González, A. (2010). Reseña de "El derecho a la autodeterminación informativa" de Murillo de la Cueva, Pablo Lucas y Piñar Mañas, José Luis. Boletín Mexicano de Derecho Comparado, XLIII (127), 421-424. citar dentro de párrafo apa 7ma edición.

García-Murillo, J. G. (2019). Turbulence of Personal Data, between Public and Private. Política, Globalidad y Ciudadanía, 4(7), 68. https://doi.org/10.29105/pgc4.7-5

Garhart, N. (2020). Data Scraping Under the Revised CCPA Regulations, JDSUPRA. https://www.jdsupra.com/legalnews/data-scraping-under-the-revised-ccpa-43249/.

Gesetz- und Verordnungsblatt (1970). Datenschutzgesetz. https://protecciondata.es/wp-content/uploads/2019/10/00041.pdf

Gil González, E. (2016). Big data, privacidad y protección de datos. España: Agencia Española de Protección de Datos.

González, Roberto. (1990). El deber de respeto a la intimidad, Pamplona, Pág. 185.

Guía Legislativa sobre la Privacidad y la Protección de Datos Personales en las Américas. (2015).

Gupta, S, Kaiser, G, Neistadt, D, & Grimm, P. (2003). DOM-based content extraction of HTML documents. In Proceedings of the 12th international conference on World Wide Web (WWW '03). *Association for Computing Machinery, New York, Estados Unidos,* 207–214. https://doi.org/10.1145/775152.775182

Grupo de Trabajo del Artículo 29. (2007). *Opinion 4/2007 on the concept of personal data*.

Grupo de Trabajo del Artículo 29. (2009). Dictamen 5/2009 del Grupo de Trabajo del Artículo 29 (pp. 4-5).

Harper, D. (s.f.). Data. In *Online Etymology Dictionary*. Recuperado de https://www.etymonline.com/word/data#etymonline_v_782

Herrán, Ana (1998). La violación de la intimidad en la protección de datos personales, Dykinson SL, Madrid, 1998, p.2.

Herrán, Ana (2002). El Derecho a la intimidad en la nueva ley orgánica de protección de datos personales. Madrid, p. 25.

Hines v. Overstock.com, 668 F. Supp. 2d 362, 366-67 (E.D.N.Y. 2009).

Hohfeld, W. (1968). Conceptos jurídicos fundamentales (Traducción de Some fundamental legal conceptions as applied to judicial reasoning, Yale Law Journal, 1913). Buenos Aires: Centro Editor de América Latina.

Hartzog, W. (2019). The Public Information Fallacy, 99 BOS. L.

H.R. 8152, 117th Cong. (2022). American Data Privacy and Protection Act. https://www.congress.gov/bill/117th-congress/house-bill/8152

Hustinx, P. (2009). Privacy by Design: Delivering the Promises. European Data Protection Supervisor. Recuperado de https://edps.europa.eu/sites/edp/files/publication/09-11-02_madrid_privacybydesign_en.pdf

Import.IO, Enterprise scale eCommerce data to drive growth - Import.io (2022). [Online]. https://www.import.io/

Information Commissioner's Office. (2022). Monetary Penalty Notice. https://ico.org.uk/media/action-weve-taken/mpns/4020436/clearview-ai-inc-mpn-20220518.pdf

Information Commissioner's Office. (2022). ENFORCEMENT POWERS OF THE INFORMATION COMMISSIONER. https://ico.org.uk/media/action-weve-taken/enforcement-notices/4020437/clearview-ai-inc-en-20220518.pdf

Information Commissioner's Office. (2023). Joint statement on data scraping and the protection of privacy. https://ico.org.uk/media/about-the-ico/documents/4026232/joint-statement-data-scraping-202308.pdf

Information Commissioner's Office. (2024). Generative AI first call for evidence: The lawful basis for web scraping to train generative AI models. https://ico.org.uk/about-the-ico/what-we-

do/our-work-on-artificial-intelligence/generative-ai-first-call-for-evidence/

Istituto di Studi Giuridici Internazionali (2010). L'accesso ad Internet: un nuovo diritto fondamentale? (2010). https://www.treccani.it/magazine/diritto/approfondimenti/diritto_internazionale_e_comparato/2_Pisa_internet.html

Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales. (2018). *Recomendaciones para mantener segura la privacidad y datos personales en el entorno digital*. Recuperado de https://inicio.inai.org.mx/Guias/5RecomendacionesPDP_Web.pdf

Internet Society (s.f.). Breve historia de internet. https://www.internetsociety.org/es/internet/history-internet/brief-history-internet/.

Internet Society. (2019). Internet Society Global Internet Report: Consolidation in the Internet Economy. *Isoc*, 75. https://future.internetsociety.org/2019/

Internet Society (2022). Página 1. https://www.internetsociety.org/wp-content/uploads/2022/07/2022-Internet-Ecosystem-EN.pdf.

Ishida, T., Sasaki, Y., & Fukuhara, Y. (1991). Use of procedural programming languages for controlling production systems. In *Proceedings of the Seventh IEEE Conference on Artificial Intelligence Applications.*

ISO/IEC (2017), Information technology -- Cloud computing -- Interoperability and portability. http://www.iso.org/standard/66639.html.

ISO/IEC (2018), Privacy enhancing data de-identification terminology and classification of techniques. http://www.iso.org/standard/69373.html.

Jarmul, K., & Lawson, R. (2017). *Python web scraping - Second edition* (2nd ed.). Packt Publishing. https://www.perlego.com/book/527121/python-web-scraping-second-edition-pdf

Jervis, P (2006), La Regulación del mercado de datos personales en Chile. Tesis, p 48.

Khan, Lina; Pozen, D. (2019). A Skeptical View of Information Fiduciaries.

Khan, F.Q., Tsaramirsis, G., Ullah, N., Nazmudeen, M.S., Jan, S., & Ahmad, A. (2020). Smart algorithmic based web crawling and scraping with template autoupdate capabilities. *Concurrency and Computation: Practice and Experience, 33*. https://doi.org/10.1002/cpe.6042

Khder, M. A. (2021). Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing and its Applications*. https://ijasca.zuj.edu.jo/PapersUploaded/2021.3.11.pdf

Kinsta. (2022). ¿Qué Es el Web Scraping? Cómo Extraer Legalmente el Contenido de la Web. https://kinsta.com/es/base-de-conocimiento/que-es-web-scraping/

Kitchin R (2014). The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences. SAGE Publications.

Kobrata, D., & Atika, R. (2021). The Privacy, Data Protection and Cybersecurity Law Review: Indonesia. The Law Reviews. https://thelawreviews.co.uk/title/the-privacy-data-protection-and-cybersecurity-law-review/indonesia

Korff, D., & Georges, M. (2015). Passenger Name Records, data mining & data protection: the need for strong safeguards. June.

Lawson, R. (2015). Web Scraping with Python: Successfully scrape data from any website with the power of Python (pp. 1–174). Packt Publishing.

Lee, M. H., Yun, J. H. J., Pyka, A., Won, D. K., Kodama, F., Schiuma, G., Park, H. S., Jeon, J., Park, K. B., Jung, K. H., Yan, M. R., Lee, S. Y., & Zhao, X. (2018). How to respond to the Fourth Industrial Revolution, or the second information technology revolution? Dynamic new combinations between technology, market, and society through open innovation. *Journal of Open Innovation: Technology, Market, and Complexity*, *4*(3). https://doi.org/10.3390/joitmc4030021

León, L. (2006). Derechos de la personalidad y medios de comunicación [Tesis de doctorado, Scuola S. Anna di Pisa].

León, L. (2011). Manipulación de información personal y derechos fundamentales. Crítica del proyecto de "ley de protección de datos personales". https://www.academia.edu/713133/Manipulaci%C3%B3n_de_informaci%C3%B3n_personal_y_derechos_fundamentales_Cr%C3%ADtica_del_proyecto_de_ley_de_protecci%C3%B3n_de_datos_personales

Ley de Fraude y Abuso Informático - Código 18 USC 1030 (1986)

Ley de Privacidad del Consumidor de California (2018)

Ley de la Privacidad de Información Biométrica de Illinois (2008)

Ley de Derechos de Privacidad de California (CPRA) (2020)

Ley de Protección de Datos Personales, Ley No. 25.326, Argentina (2000).

Ley de Telecomunicaciones-Telemedios-Protección de Datos (Telekommunikation-Telemedien-Datenschutzgesetz), 2021.

Ley 1266 de 2008, Ley estudiada por la Corte Constitucional mediante Sentencia C-1011 de 2008 por la cual se dictan las disposiciones generales del hábeas data y se regula el manejo de la información contenida en bases de datos personales, en especial la financiera, crediticia,

comercial, de servicios y la proveniente de terceros países y se dictan otras disposiciones. (2008). https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=34488#:~:text=por%20la%20cual%20se%20dictan,y%20se%20dictan%20otras%20disposiciones

Ley 1266 de 2008. Ley hábeas data y manejo de la información contenida en bases de datos personales. (2008). https://www.bogotajuridica.gov.co/sisjur/normas/Norma1.jsp?i=34488

Ley 1581 de 2012. Por la cual se dictan disposiciones generales para la protección de datos personales (2012).

Ley Federal de Protección de Datos - *Bundesdatenschutzgesetz* (1977)

Ley Federal Alemana de Protección de Datos - Bundesdatenschutzgesetz (2018)

Ley No. 29733, Ley de Protección de Datos Personales. Diario Oficial El Peruano (2011). https://spij.minjus.gob.pe/spij-ext-web/#/detallenorma/H1034642

Ley No. 30254, Ley de Promoción para el Uso Seguro y Responsable de las Tecnologías de la Información y las Comunicaciones por Niños, Niñas y Adolescentes. Diario Oficial El Peruano (2014). https://www.gob.pe/institucion/congreso-de-la-republica/normas-legales/292146-30254

Ley No. 31307, Nuevo Código Procesal Constitucional. Diario Oficial El Peruano (2021). https://cdn.gacetajuridica.com.pe/laley/nuevo-codigo-procesal-constitucional-ley-no-31307-1975873-2%20A.pdf

Ley No. 31814, Ley que promueve el uso de la inteligencia artificial en favor del desarrollo económico y social del país. (13 de junio de 2023). https://busquedas.elperuano.pe/dispositivo/NL/2192926-1

Ley Orgánica 5/1992, de 29 de octubre, de regulación del Tratamiento Automatizado de los Datos de Carácter Personal. (1992). https://www.boe.es/buscar/doc.php?id=BOE-A-1992-24189

Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.

Linares, S. (2019). El contenido constitucional del derecho fundamental a la autodeterminación informativa en el Derecho Constitucional peruano (Tesis para optar el título de Abogado).

Universidad de Piura. Facultad de Derecho. Programa Académico de Derecho. Lima, Perú. https://pirhua.udep.edu.pe/items/19a5b32a-cf46-4c1c-a1c4-cb2587909770

LinkedIn Ireland Unlimited Company (2023). Resolución No. 71406.

Lucas Verdú, P. (1995). Prólogo. En La configuración constitucional del derecho a la intimidad (p. 22). Madrid.

Luna Cervantes, E. J. (2021). Preguntas y respuestas varias sobre la protección de datos personales en el Perú. Advocatus, (039), 253-264. https://doi.org/10.26439/advocatus2021.n39.5133

Marriott International, Inc., Customer Data Security Breach Litigation. (2023). MDL No. 19-md-2879. https://caselaw.findlaw.com/court/us-dis-crt-d-mar-gre-div/115563914.html

Martini, B. (2015). The Data Revolution. Big Data, Open Data, Data Infrastructures and Their Consequences. *Regional Studies*. https://doi.org/10.1080/00343404.2015.1107987

Millard, C., & Hon, W.K. (2011). Defining 'Personal data' in E-social Science. *Information, Communication & Society*, *15*, 66 - 84.

Ministerio de Justicia y Derechos Humanos. (2020). Metodología para el Cálculo de las Multas en materia de Protección de Datos Personales. https://bnl.minjus.gob.pe/bnl/PreviewSecure?dw=1&uuid=cF%2BpTGrDYHub93S53O%2BbWh0yvf67sOA9nyXrMpk%2Bffoa%2BYFZqHH%2BjQ%3D%3D

Ministerio de Justicia y Derechos Humanos del Perú. (2023). Autoridad Nacional de Protección de Datos Personales impuso multas por más de S/ 7.6 millones durante el 2023. https://www.gob.pe/institucion/minjus/noticias/893316-autoridad-nacional-de-proteccion-de-datos-personales-impuso-multas-por-mas-de-s-7-6-millones-durante-el-2023

Ministerio de Justicia y Derechos Humanos. (2024). Proyecto de Reglamento de la Ley No. 29733, Ley de Protección de Datos Personales.

Ministerio de Justicia y Derechos Humanos. (2014). Oficio No. 569-2014-JUS/DGPDP.

Ministerio de Justicia y Derechos Humanos. (2021). Resolución No. 1264-2021-JUS/DGTAIPD-DPDP.

Ministerio de Justicia y Derechos Humanos. (2021). Resolución No. 3442-2021-JUS/DGTAIPD-DPDP.

Ministerio de Justicia y Derechos Humanos. (2018). Resolución No. 3551-2018-JUS/DGTAIPD-DPDP, foja 35-38.

Ministerio de Justicia y Derechos Humanos. (2018). Opinión Consultiva No. 749-2018-DGTAIPD.

Ministerio de Justicia y Derechos Humanos. (2019). Resolución Directoral No. 1623-2019-JUS/DGTAIPD-DPDP.

Ministerio de Justicia y Derechos Humanos. (2023). Opinión Consultiva No. 020-2023-DGTAIPD.

Ministerio de Justicia, Seguridad y Derechos Humanos. Argentina (2009). Nota JGM N° 738/09: Dictamen DNPDP N° /09. Recuperado de https://www.argentina.gob.ar/sites/default/files/d2009_32.pdf

Muñoz Vela, José Manuel (2022). Retos, riesgos, responsabilidad y regulación de la inteligencia artificial. Aranzadi. https://www.dykinson.com/autores/munoz-vela-jose-manuel/47512/

Murphy, G. (2019). Beyond Integrated Development Environments: Adding Context to Software Development. En *Proceedings of the 2019 IEEE/ACM 41st International Conference on Software Engineering: New Ideas and Emerging Results* (pp. 73-76). https://doi.org/10.1109/ICSE-NIER.2019.00027

Murillo, P. L., & Piñar, J. (1990). El derecho a la autodeterminación informativa. Madrid: TECNOS.

Murillo de la Cueva, P. L. (1993). Informática y protección de datos personales. Estudio sobre la Ley Orgánica 5/1992, de regulación del tratamiento automatizado de los datos de carácter personal. Cuadernos y Debates. Madrid: Centro de Estudios Constitucionales.

Mrd0x. (s.f.). *EvilSelenium*. GitHub. https://github.com/mrd0x/EvilSelenium

Muñoz Vela, J. M. (2022). Retos, riesgos, responsabilidad y regulación de la inteligencia artificial: Un enfoque de seguridad física, lógica, moral y jurídica. España: Aranzadi/Civitas.

Naciones Unidas. (1966). Pacto Internacional de Derechos Civiles y Políticos. https://www.ohchr.org/es/professionalinterest/pages/ccpr.aspx

Narayanan, A. and V. Shmatikov (2006), "How To Break Anonymity of the Netflix Prize Dataset", CoRR abs/cs/0610105. http://arxiv.org/abs/cs/0610105.

Nguyen v. Barnes & Noble Inc., 763 F.3d 1171, 1176 (9th Cir. 2014).

Norma de protección de la privacidad en línea de los niños (1998)

OCDE sobre protección de la privacidad y flujos transfronterizos de datos personales.

OdiseIA & Observatorio del Impacto Social y Ético de la Inteligencia Artificial. (2022). Guía de buenas prácticas para el uso de la inteligencia artificial ética.

Organización de los Estados Americanos. (2021). *Protección de datos personales: Principios actualizados 2021*. https://www.oas.org/es/sla/cji/docs/Publicacion_Proteccion_Datos_Personales_Principios_Actualizados_2021.pdf

Organización de los Estados Americanos. (1948). Declaración Americana de los Derechos y Deberes del Hombre. https://www.oas.org/es/cidh/mandato/basicos/Declaracion%20Americana%20de%20los%20Derechos%20y%20Deberes%20del%20Hombre.asp

Organización de los Estados Americanos. (1969). Convención Americana sobre Derechos Humanos - Pacto de San José. https://www.oas.org/es/cidh/mandato/basicos/convencion%20Americana%20derechos%20humanos.asp

Organización de los Estados Americanos. (1980.). *Directrices de la OCDE sobre privacidad.* https://www.oas.org/es/sla/ddi/docs/directrices_ocde_privacidad.pdf

Organización para la Cooperación y el Desarrollo Económico. (2013). Directrices de la Osiptel. (2012). *Texto Único Ordenado de las Condiciones de Uso de los Servicios Públicos de Telecomunicaciones*. Resolución de Consejo Directivo No. 138-2012-CD/Osiptel.

Organización para la Cooperación y el Desarrollo Económico. (2014), Summary of OECD Expert Roundtable: "Protecting Privacy in a Data-driven Economy: Taking Stock of Current Thinking", OECD, Paris, http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=dsti/iccp/reg%282014%293&doclanguage=en.

Organización para la Cooperación y el Desarrollo Económico. (2019). Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies. OECD Publishing. https://doi.org/10.1787/276aaca8-en

Organización para la Cooperación y el Desarrollo Económico.. (2023). Recommendation of the Council on Artificial Intelligence. https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449

Pagallo, U., & Ciani, J. (September 28, 2023), Anatomy of Web Data Scraping: Ethics, Standards, and the Troubles of the Law. http://dx.doi.org/10.2139/ssrn.4707651

Parks, A. M. (2022). Unfair Collection: Reclaiming Control of Publicly Available Personal Information from Data Scrapers. Michigan Law Review, 120(5), 913–945. https://doi.org/10.36644/mlr.120.5.unfair

Parlamento Europeo y Consejo. (2012). Propuesta de Reglamento general de protección de datos del Parlamento Europeo y del Consejo. https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX%3A52012PC0011

Parlamento Europeo y Consejo (2019). Directiva (UE) 2019/790, de 17 de abril de 2019, sobre los derechos de autor y derechos afines en el mercado único digital y por la que se modifican las Directivas 96/9/CE y 2001/29/CE. https://eur-lex.europa.eu/legal-content/ES/ALL/?uri=CELEX%3A32019L0790

Pedrero Sánchez, JF. (2023). Desarrollo de procedimientos de valoración funcional mediante sensores portables. Estudios de aplicación en enfermedades neurodegenerativas: Parkinson y Alzheimer [Tesis doctoral]. Universitat Politècnica de València. https://doi.org/10.4995/Thesis/10251/202450

Peyrano, G. (2007). "El derecho a la intimidad informática: garantía de la privacidad personal en los entornos virtuales de las comunicaciones electrónicas". Jurisprudencia Argentina. 2007-IV

PM v. OpenAI LP, No. 3:23-cv-03199 (2023). https://www.bloomberglaw.com/public/desktop/document/PMetalvOPENAILPetalDocketNo323cv03199NDCalJun282023CourtDocket/1?doc_id=X1Q5O7KNE0B9N58DMJL3VN7K9SN.

Poder Ejecutivo Nacional. (2009). *Decreto 32/2009: Policía de Seguridad Aeroportuaria - Confirmación de personal*. Fecha de sanción 26-01-2009. Publicada en el Boletín Nacional del 05-Feb-2009. Recuperado de https://www.argentina.gob.ar/normativa/nacional/decreto-32-

2009-150107

Post, R. C. (2018). Data privacy and dignitary privacy: Google Spain, the right to be forgotten, and the construction of the public sphere. Duke Law Journal, 67(5), 981–1072. https://doi.org/10.2139/ssrn.2953468

Presidencia del Consejo de Ministros. (2021). *Estrategia Nacional de Inteligencia Artificial (ENIA)*. https://www.gob.pe/institucion/pcm/informes-publicaciones/1929011-estrategia-nacional-de-inteligencia-artificial

Prego, J. A. (2017). *La transparencia como elemento de apoyo al consentimiento en materia de protección de datos.* Universidad Carlos III de Madrid. https://hdl.handle.net/10016/26447

Prevos, P. (2019). Principles of Strategic Data Science: Creating Value From Data, Big and Small. Packt Publishing.

Proyecto de ley No. 375 de la Asamblea (2018). Privacy: personal information: businesses. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375

Proposición 24, la C. (2023). https://vig.cdn.sos.ca.gov/2020/general/pdf/topl-prop24.pdf

GPDP. (2022). Ordinanza ingiunzione nei confronti di Clearview AI - 10 febbraio 2022. https://www.gpdp.it/web/guest/home/docweb/-/docweb-display/docweb/9751362.

GPDP. (2023). Provvedimento del 17 maggio 2023 [9903067]. https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9903067

GPDP. (2023). Medida de 21 de diciembre de 2023 [9972593]. https://www.gpdp.it/web/guest/home/docweb/-/docweb-display/docweb/9972153

GPDP. (2024). Web scraping ed intelligenza artificiale generativa - nota informativa e possibili azioni di contrasto.pdf. https://www.gpdp.it/web/guest/home/docweb/-/docweb-display/docweb/10020334

Puccinelli, O. R. (1997). Tipos y subtipos de hábeas data en el Derecho Constitucional Latinoamericano. La Ley, - D – 222.

Quiroga, J. (2013). Apuntes sobre la protección de datos personales. A propósito del nuevo reglamento. *Revista Actualidad Jurídica (Tomo Número 233).*

Quiroz Papa de García, R. (2016). El hábeas data, protección al derecho a la información y a la autodeterminación informativa. Letras (Lima), 87(126), 23-27. http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S2071-50722016000200002&lng=es&tlng=es

Ramos-Andres, Otoniel & Gutierrez-Pulido, Rafael & Herrera-Morales, Jose-Roman & De la Torre-Gea, Guillermo. (2019). Recuperación de metadatos e indicadores de impacto para

publicaciones científicas mediante servicios de Google académico.

Ramsey, N. (2022). *Programming Languages: Build, Prove, and Compare*. Cambridge: Cambridge University Press.

Real Academia Española. (s.f.). *Información*. Recuperado el 26 de junio de 2024, de https://dle.rae.es/informaci%C3%B3n

Real Decreto 1720/2007, de 21 de diciembre, por el que se aprueba el Reglamento de desarrollo de la Ley Orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal (LOPD). Boletín Oficial del Estado.

Rebollo, L. (2003). Balance constitucional: artículo 18.4 CE. En datospersonales.org, Revista de la Agencia de Protección de Datos de la Comunidad de Madrid, (6), 2.

Red Iberoamericana de Protección de Datos (2021). Recomendaciones para el tratamiento de datos personales en servicios de nube. https://www.redipd.org/sites/default/files/2021-06/recomendaciones-tratamiento-datos-personales-servicios-nube.pdf

Red Iberoamericana de Protección de Datos. (2007). Directrices de la Red Iberoamericana de Protección de Datos.

Red Iberoamericana de Protección de Datos. (2019). Orientaciones Específicas para el Cumplimiento de los Principios y Derechos que Rigen la Protección de los Datos Personales en los Proyectos de Inteligencia Artificial. https://www.redipd.org/sites/default/files/2020-02/guia-orientaciones-específicas-proteccion-datos-ia.pdf%0Ahttps://www.redipd.org/sites/default/files/2020-02/guide-specific-guidelines-ai-projects.pdf%0Ahttps://www.redipd.org/es/documentos/guias

Register.com, Inc. v. Verio, Inc., No. 02-9531 (2d Cir. Jan. 23, 2004).

Reglamento de la Ley No. 29733. El Peruano [Decreto Supremo 003-2013-JUS], de 22 de marzo de 2013. http://www.minjus.gob.pe/wpcontent/uploads/2013/04/DS-3-2013-JUS.REGLAMENTO.LPDP (2011) .pdf.

Resolución Directoral No. 1623-2019-JUS/DGTAIPD-DPDP. Ministerio de Justicia. (12 de junio de 2019). https://cdn.www.gob.pe/uploads/document/file/1380428/1623-2019-JUSDGTAIPD-DPDP.pdf.pdf?v=1602959153

Resolución Directoral No. 1085-2022-JUS/DGTAIPD-DPDP. Expediente N.° 198-2021-PTT. https://cdn.www.gob.pe/uploads/document/file/3510295/EXP.%20198-2021%20-%20RD%201085-2022-DPDP.pdf.pdf

Resolución No. 58834 (2023), Superintendente Delegado de Protección de Datos, Ministerio de Comercio, Industria y Turismo, Superintendencia de Industria y Comercio. Recuperado de: https://www.sic.gov.co/sites/default/files/boletin-juridico/18-153189.pdf

Remolina Angarita, N. (E.d) (2015). Recolección internacional de datos personales: un reto del mundo post-Internet. Agencia Española de Protección de Datos (AEPD). Boletín Oficial del Estado. ISBN 987-8434021969. Recuperado de https://www.google.com.pe/books/edition/Recolecci%C3%B3n_internacional_de_datos_pers/BaQFEAAAQBAJ?hl=es&gbpv=0

Reynolds, E. (29 de junio de 2020). Psychologists are mining social media posts for mental health research — but many users have concerns. *British Psychological Society*. Recuperado de https://digest.bps.org.uk/2020/06/29/psychologists-are-mining-social-media-posts-for-mental-healthresearch-but-many-users-have-concerns/ [https://perma.cc/3USB-YLA4]

Riley, K. (2018). Data Scraping as a Cause of Action: Limiting Use of the CFAA and Trespass in Online Copying Cases, 29 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. p- 245, 272–76.

Rivera Llano, A. (1984). La protección de la intimidad y el honor y la informática. En J. E. Valencia (Ed.), *Estudios Penales. Homenaje al profesor Luis Carlos Pérez* (pp. 171). TEMIS.

Rodríguez Pardo, J. (2011). Derecho de la información.: Una perspectiva comparada de España e Iberoamérica. España: Editorial Dykinson, S.L.

Rodrigo, J. J., & Rosales, J. J. (2017). Extracción de conocimiento mediante la aplicación de inteligencia artificial a la información espacial. Spatial Artificial Inteligence Knowledge

Rojas, Marcela. (2015). Evolución del derecho de protección de datos personales en Colombia respecto a estándares internacionales. Universidad Católica de Colombia. Número: Vol. 8 Núm. 1 (2014): Enero – Junio. https://novumjus.ucatolica.edu.co/article/view/652

Russell, S. J., & Norvig, P. (2004). *Inteligencia artificial: Un enfoque moderno* (2ª ed.). Madrid, España: Pearson Educación S.A. https://luismejias21.files.wordpress.com/2017/09/inteligencia-artificial-un-enfoque-moderno-stuart-j-russell.pdf

Rojas Bejarano, M. (2014). Evolución del derecho de protección de datos personales en Colombia. Novum Jus, 8, 107–139. https://editorial.ucatolica.edu.co/ojsucatolica/revistas_ucatolica/index.php/Juridica/article/viewFile/652/670

Rosenberg, D. (2013) 'Data before the fact', in L. Gitelman (ed.), 'Raw Data' is an Oxymoron. MIT Press, Cambridge, MA, pp. 15–40.

Ryanair Ltd v Billigfluege.de GmbH and Others, Corte Suprema. (2015). https://ie.vlex.com/vid/ryanair-ltd-v-billigfluege-793392113

Sánchez Perez, A. (2022). Generación Datos Sintéticos [Trabajo de Fin de Grado, Universidad Politécnica de Madrid].

Santamaría Ramos, F. J. (2011). El encargado independiente. Figura clave para un nuevo Derecho de protección de datos. La Ley grupo Wolters Kluwer. Madrid.

Scroxton, A. (2020). Social media data leak highlights murky world of data scraping. *Computer Weekly*. https://www.computerweekly.com/news/252487895/Social-media-data-leak-highlights-murky-world-of-datascraping [https://perma.cc/52W9-R5UM]

Schwartz, P. M. (2004). Property, Privacy, and Personal Data. Harvard Law Review, 117(8), 2056-2128. Recuperado de https://www.jstor.org/stable/4093335

Secretaría de Gobierno y Transformación Digital (2024). Resolución No. 001-2024-PCM/SGTD.

Secretaría de Gobierno y Transformación Digital. (2018). Lineamientos para la Formulación del Plan de Gobierno Digital - PGD. http://spij.minjus.gob.pe/Graficos/Peru/2018/Diciembre/22/RSGD-005-2018-PCM-SEGDI.pdf

SeleniumHQ. (s.f.). *SeleniumHQ GitHub Page*. GitHub. Recuperado de https://github.com/SeleniumHQ

Sierra, J.L., Navarro, A., Fernandez-Manjon, B., & Fernández-Valmayor, A. (2005). Incremental definition and operationalization of domain-specific markup languages in ADDS. *ACM SIGPLAN Notices*, 40, 28-37.

Silva, C.P., Feitosa, E.L., & Garcia, V.C. (2017). Uma Avaliação das Prevenções de Phishing em Navegadores Web. *A*nais do XVII Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais (SBSeg 2017).

Sinche, J. U., & Torres, J. C. (2021). Análisis de sentimientos en los mensajes recibidos en el entorno virtual de aprendizaje de la modalidad abierta y a distancia de la UTPL. RISTI - Revista Ibérica de Sistemas y Tecnologías de la Información, 41, 98- 113. http://www.risti.xyz.

Singh, B.P., & Singh, H.K. (2010). Web Data Mining research: A survey. 2010 IEEE International Conference on Computational Intelligence and Computing Research, 1-10.

Solove, D. J. (2024). Data Is What Data Does: Regulating Based on Harm and Risk Instead of Sensitive Data. 118 Northwestern University Law Review 1081. GWU Legal Studies Research Paper No. 2023-22. GWU Law School Public Law Research Paper No. 2023-22. https://ssrn.com/abstract=4322198

Solove, D. J. (2006). A Taxonomy of Privacy. *University of Pennsylvania Law Review, 154*(3), 477. Recuperado de https://scholarship.law.upenn.edu/penn_law_review/vol154/iss3/1

Specht v. Netscape Commc'ns, 306 F.3d 17, 22 n.4 (2d Cir. 2002).

Stanford University. (2023). *Artificial Intelligence Index Report 2023*. Retrieved from https://aiindex.stanford.edu/report/

Subirats Maté, L., & Calvo González, M. C. (2019). *Web scraping* (p. 66). https://openaccess.uoc.edu/bitstream/10609/147437/1/WebScraping_Modulo1_WebScraping.pdf

Sucursal, D. B. A. G., Website, M., Sucursal, D. B. A. G., Services, C., E-mail, B., Officer, D. P., & Services, C. (2018). Data protection information under the EU General Data Protection Regulation in Spain. 6(1), 6–8.

Susser, D. (2019). Notice After Notice-and-Consent: Why Privacy Disclosures Are Valuable Even If Consent Frameworks Aren't.

Superintendencia de Industria y Comercio (SIC). Resolución No. 63771, 2022. Ministerio de Comercio, Industria y Turismo. https://www.sic.gov.co/sites/default/files/documentos/102022/RE63771-2022.pdf

Superintendencia de Industria y Comercio (SIC). Resolución No. 58834 de 2023. Ministerio de Comercio, Industria y Turismo. https://www.ambitojuridico.com/sites/default/files/2024-01/RES-58834-2023.pdf

Sweeney, L. (2000). Simple Demographics Often Identify People Uniquely (Carnegie Mellon

Univ. Data Priv. Working Paper, Paper No. 3). https://dataprivacylab.org/projects/identifiability/paper1.pdf

S. 2361, 100th Cong. (1988). Video Privacy Protection Act. https://www.congress.gov/bill/100th-congress/senate-bill/2361/text

Teledienstedatenschutzgesetz (1996), BGBl.

Torres, J. (2018). *DEEP LEARNING: Introducción práctica con Keras*. España: Lulu.com. Recuperado de https://books.google.com.pe/books/about/DEEP_LEARNING_Introducci_n_pr_ctica_con.html?id=ju1mDwAAQBAJ&redir_esc=y

Tribunal Constitucional Español. (1993). Sentencia 254/1993, de 20 de julio.

Tribunal Constitucional. (1998). Sentencia recaída en el Expediente No. 666-96-HD/TC de fecha 02 de abril de 1998. https://www.tc.gob.pe/jurisprudencia/1998/00666-1996-HD.html

Tribunal Constitucional Español. (2000). Sentencia 290/2000 de 30 de noviembre.

Tribunal Constitucional Español. (2000). Sentencia 292/2000 de 30 de noviembre.

Tribunal Constitucional. (2003). Sentencia recaída en el Expediente No. 01797-2002-HD/TC de fecha 29 de enero de 2003. https://www.tc.gob.pe/jurisprudencia/2003/01797-2002-HD.html

Tribunal Constitucional. (2004). Sentencia recaída en el Expediente No. 1219-2003-HD de fecha 21 de enero de 2004. https://tc.gob.pe/jurisprudencia/2004/01219-2003-HD.pdf

Tribunal Constitucional. (2005). Sentencia recaída en el Expediente No. 0072-2004-AA/TC de fecha 7 de abril de 2005. https://tc.gob.pe/jurisprudencia/2005/00072-2004-AA.pdf

Tribunal Constitucional (2005). Sentencia recaída en el Expediente No. 06712-2005-HC/TC de fecha 17 de octubre de 2005. https://www.tc.gob.pe/jurisprudencia/2006/06712-2005-HC.pdf

Tribunal Constitucional. (2006). Sentencia recaída en el Expediente No. 047-2004-AI/TC de fecha 24 de abril de 2006. https://www.tc.gob.pe/jurisprudencia/2006/00047-2004-AI.html

Tribunal Constitucional. (2007). Sentencia recaída en el Expediente No. 00009-2007-PI/TC de fecha 29 de agosto de 2007. https://www.tc.gob.pe/jurisprudencia/2007/00009-2007-AI%2000010-2007-AI.html

Tribunal Constitucional. (2007). Sentencia recaída en el Expediente No. 04739-2007-PHD/TC de fecha 15 de octubre de 2007. https://tc.gob.pe/jurisprudencia/2008/04739-2007-HD.pdf

Tribunal Constitucional. (2007). Sentencia recaída en el Expediente No. 6164-2007-PHD/TC de fecha 21 de diciembre de 2007. https://www.tc.gob.pe/jurisprudencia/2008/06164-2007-HD.pdf

Tribunal Constitucional (2011). Sentencia recaída en el Expediente No. 02838-2009-PHD/TC de fecha 31 de enero de 2011. https://www.tc.gob.pe/jurisprudencia/2011/02838-2009-HD.html

Tribunal Constitucional (2011). Sentencia recaída en el Expediente No. 04387-2011-PHD/TC de fecha 29 de agosto de 2013. https://www.tc.gob.pe/jurisprudencia/2013/04387-2011-HD.html

Tribunal Constitucional (2014). Sentencia recaída en el Expediente No. 01839-2012-PHD/TC de fecha 09 de diciembre de 2014. https://www.tc.gob.pe/jurisprudencia/2016/01839-2012-HD.pdf

Tribunal Constitucional (2017). Sentencia recaída en el Exp. 00442-2017-PA/TC

Tribunal Constitucional. (2021). Pleno Sentencia 404/2021 de fecha 18 de marzo de 2021. https://www.tc.gob.pe/jurisprudencia/2021/04038-2019-HC.pdf

Tribunal Constitucional (2022). Sentencia recaída en el Exp. 01163-2022-PHD/TC.

Tribunal Constitucional. (2022). Sentencia recaída en el Expediente No. 02839-2021-PHD/TC de fecha 22 día de agosto de 2022. https://www.tc.gob.pe/jurisprudencia/2021/04038-2019-HC.pdf

Tribunal Constitucional. (2022). Sentencia recaída en el Expediente No. 3041-2021-PHD/TC de fecha 17 de junio de 2022. https://tc.gob.pe/jurisprudencia/2022/03041-2021-HD.pdf

Tribunal Constitucional. (2022). Sentencia recaída en el Expediente No. 3100-2021-PHD/TC de fecha 28 de febrero de 2022. https://img.lpderecho.pe/wp-content/uploads/2022/04/Expediente-03100-2021-PHD-TC-LPDerecho.pdf

Tribunal Constitucional. (2022). Sentencia recaída en el Expediente No. 5060-2009-PHD/TC.

Tribunal de Justicia de la Unión Europea. (2011). Caso C-70/10. https://curia.europa.eu/juris/liste.jsf?language=en&num=C-70/10

Tribunal de Justicia de la Unión Europea. (2015). Court of Justice of the European Union: Press

Release No 117/15. https://curia.europa.eu/jcms/upload/docs/application/pdf/2015-10/cp150117en.pdf

Tribunal de Justicia de la Unión Europea. (2016). Caso C-582/14. https://curia.europa.eu/juris/liste.jsf?num=C-582/14

Tribunal del Noveno Circuito de los Estados Unidos. (2019). Expediente N° 17-16783 del 9 de septiembre de 2019. [Caso LinkedIn contra HiQ Labs]. Estados Unidos.

Tribunal Federal de Justicia Alemán (2014). Sentencia de 30 de abril de 2014 - I ZR 224/12 - Corretaje de vuelos en Internet. https://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/document.py?Gericht=bgh&Art=en&Datum=Aktuell&nr=67581&linked=pm

Truong, L., & Hanrahan, P. (2019). A Golden Age of Hardware Description Languages: Applying Programming Language Techniques to Improve Design Productivity. En *3rd Summit on Advances in Programming Languages (SNAPL 2019)*. Leibniz International Proceedings in Informatics (LIPIcs), Volume 136, pp. 7:1-7:21, Schloss Dagstuhl – Leibniz-Zentrum für Informatik. https://doi.org/10.4230/LIPIcs.SNAPL.2019.7

Unión Europea. (2000). Carta de los Derechos Fundamentales de la Unión Europea.

Utah Code Annotated. (2023). § 13-61-101(32). West.

Unión Europea. Tribunal de Justicia de la Unión Europea. (2011). Sentencia de 24 de noviembre de 2011, asuntos acumulados C-468/10 y C-469/10, ECLI:EU:C:2011:777 apartados 24-49.

Unión Europea. (2021). Artificial Intelligence Act. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206

UK. (2018). Data Protection Act 2018. https://www.legislation.gov.uk/ukpga/2018/12/introduction/enacted

U.S. Department of Justice. Privacy Act of 1974. https://www.justice.gov/opcl/privacy-act-1974

Vásquez, P. C., & Kramcsák, P. T. (2019). Legitimate interest and personal data processing: Comparative background and Chilean regulation. Revista Chilena de Derecho y Tecnología,

8(1), 69–107. https://doi.org/10.5354/0719-2584.2019.52915

Valverde González, V. L. (2019). Algoritmo para el diseño de base de datos. *Ciencia Digital*, *3*(3.2.1), 5-19. https://doi.org/10.33262/cienciadigital.v3i3.2.1.778

Vallez, N., Velasco Mata, A., Cotorro, J. J., & Deniz, Ó. (2019). ¿Es posible entrenar modelos de aprendizaje profundo con datos sintéticos? 859–865. https://doi.org/10.17979/spudc.9788497497169.859

Vanden Broucke, S., & Baesens, B. (2018). Introduction. In *Practical Web Scraping for Data Science: Best Practices and Examples with Python* (pp. 3–23). Apress. https://doi.org/10.1007/978-1-4842-3582-9_1

Vargiu, E., & Urru, M. (2013). Exploiting web scraping in a collaborative filtering-based approach to web advertising. *Artificial Intelligence Research, 2*(1), 44. https://doi.org/10.5430/air.v2n1p44

Vassio, L., Drago, I., Mellia, M., Ben Houidi, Z., & Lamali, M. L. (2018). You, the Web, and Your Device: Longitudinal Characterization of Browsing Habits. *ACM Transactions on the Web (TWEB), 12*(4). Article 24, página 30. https://doi.org/10.1145/3231466

Wacks, R. (1993). Personal Information. Privacy and the Law. Clarendon Press. Oxford.

Wardhan, H., & Madan, D.S. (2021). Study of Functioning of Selenium Testing Tool.

Weinberger, D. (2011). *Too Big to Know*. Basic Books, New York.

W3C. (2014, July 10). *Website Accessibility Conformance Evaluation Methodology (WCAG-EM) 1.0*. W3C Working Group Note. https://www.w3.org/TR/WCAG-EM/

Xiao, G. (2021). Bad Bots: Regulating the Scraping of public personal information. In Harvard Journal of Law & Technology (Volume 34, Number 2). https://jolt.law.harvard.edu/assets/articlePDFs/v34/6.-Xiao-Bad-Bots-Regulating-the-Scraping-of-Public-Personal-Information-edit.pdf

Zhang, J., Wang, Q., Yang, Q., Zhou, R., & Zhang, Y. (2018). Exploiting multi-category characteristics and unified framework to extract web content. *Data Science and Engineering*, *2*, 101-114. https://doi.org/10.1007/s41019-018-0067-3