Introducción a la Gobernanza de Internet en LAC

Módulo 6.1: Temas emergentes del ecosistema digital contemporáneo.

En esta oportunidad se abordarán tres temas que están hoy en la agenda de la gobernanza de Internet, y aun cuando no son estrictamente novedosos en términos de su creación, su avance está generando impacto en este escenario.

Estos temas sobre los que nos enfocaremos son: **Internet de las Cosas**, también conocida como la **Internet of Things, o IoT** por su acrónimo en inglés. Una segunda tecnología es blockchain y la tercera es la Inteligencia Artificial. Las tres son tecnologías de propósito general, esto es, sirven para múltiples usos y aplicaciones y en todos los casos el uso de los protocolos de Internet resulta fundamental para expandir su potencial.

Internet Society, en su reporte **"The Internet of Things: An overview"** expresa que el término Internet de las cosas generalmente se refiere a escenarios en los que la conectividad de red y la capacidad informática se extienden a objetos, sensores y elementos cotidianos que normalmente no se consideran computadoras, lo que permite que estos dispositivos generan, intercambien y consuman datos con una mínima intervención humana. Sin embargo, no existe una definición única y universal.

Internet de las cosas consiste, esencialmente, en colocar ordenadores y sensores a todo tipo de bienes, de modo que puedan recoger información de su uso y del ambiente que les rodea para transmitirla a otros bienes, también computarizados, e interactuar con ellos. Debido a la reducción de los precios de los elementos informáticos y la posibilidad de colocarlos en cualquier bien, se calcula que la cantidad de aparatos conectados a Internet podría escalar de forma exponencial en los próximos años de acuerdo con el informe de Device Democracy.

Internet de las Cosas (IoT) trae aparejadas oportunidades y riesgos al Sistema de Nombres de Dominio (DNS):

- Es una **oportunidad** porque el DNS proporciona funciones y datos que pueden ayudar a que IoT sea más seguro, estable y transparente, lo cual es crítico dada la interacción de IoT con el mundo físico, tal como se explicará en las próximas diapositivas.

- Es un **riesgo** porque varios estudios de medición sugieren que los dispositivos IoT pueden estresar el DNS, por ejemplo, debido a los complejos ataques DDoS llevados a cabo por botnets que crecen hasta cientos de miles o en el futuro millones de dispositivos IoT infectados en cuestión de horas.

  Un **ataque DDoS** es un tipo de ataque que busca que un determinado sitio web quede sin la posibilidad de seguir ofreciendo servicios a sus usuarios. Esto lo hacen solicitando un número de servicios superior al que puede resistir el servidor del sitio web atacado. Los ataques se pueden producir simultáneamente por varios servidores de cualquier parte del mundo y situaciones geográficas diferentes. El uso de cada vez más tipos diferentes de dispositivos conectados para brindar servicio al IoT implica que dichos dispositivos puedan ser utilizados para organizar estos ataques DDoS sin el conocimiento de sus usuarios, especialmente teniendo en cuenta que muchos de estos dispositivos no tienen "paneles de control" que puedan ser controlados por los usuarios. Por ejemplo, lavadoras, refrigeradores, lámparas eléctricas.

  Una **botnet** es una red de equipos /dispositivos que se prestan para enviar spam, para ataques DDos o para el alojamiento de información y actividades ilegales. Estos equipos suelen estar infectados con códigos maliciosos y responden a las órdenes de un solo atacante, que opera de forma remota.

Las implementaciones de IoT introducen nuevos requisitos de seguridad, disponibilidad y transparencia porque interactúan con el espacio físico, con frecuencia, sin intervención del ser humano o de su conocimiento.

Esta es una oportunidad para aumentar aún más el valor del DNS porque es una infraestructura generalizada que puede ayudar a cumplir estos requisitos. Específicamente, el DNS:

- Permite que los dispositivos IoT autentiquen resolutores y cifren el tráfico DNS: mediante el uso de nuevos protocolos para autenticar resolutores de dispositivos IoT y encriptar consultas, como DNS-over-HTTPS y DNS-over-TLS, que son dos nuevos protocolos que encriptan mensajes DNS entre un cliente DNS y su resolutor, ocultando así las búsquedas de dominio y respuestas de inspección en ruta y / o alteración entre el cliente y el resolutor.

- Admite la autenticidad del mensaje: mediante el uso de DNSSEC (*Domain Name System Security Extensions)* que son un conjunto de extensiones desarrolladas por la Internet Engineering Task Force (IETF) que refuerzan la autenticación en el DNSSEC mediante el uso de *firmas digitales* basadas en la *criptografía de*

clave pública. Con DNSSEC, no son las consultas y las respuestas del DNS en sí las que están criptográfc icamente firmadas, sino que los datos del DNS en sí están firmados por el propietario de los datos.

- DNSSEC se usa para detectar redireccionamientos maliciosos de dispositivos IoT, ya que, el DNSSEC proporciona integridad de mensajes en el DNS, lo que significa que los resolutores pueden detectar adversarios que realizan cambios en el contenido de los mensajes DNS. Esto es particularmente importante en IoT porque los mensajes DNS manipulados pueden redirigir los dispositivos IoT a un servicio malicioso, lo que pone en peligro la privacidad, la seguridad y el bienestar de los usuarios. Además, los usuarios pueden no ser conscientes de tales redirecciones, por ejemplo, porque los dispositivos pueden no ser interactivos y operar de forma autónoma.

- Proporciona formas adicionales para que los dispositivos IoT verifiquen la validez de los servicios a los que se conectan: También existen varios protocolos de seguridad basados en DNS que brindan a los dispositivos IoT medios adicionales para verificar la autenticidad de los certificados que reciben de los servicios después de la resolución DNS. Esto también es importante para evitar que los dispositivos IoT se comuniquen con servicios maliciosos, lo cual podría afectar la privacidad y seguridad de los usuarios.

- Ofrece una oportunidad para que los proveedores de registro de nombres de dominio brinden servicios de seguridad adicionales para implementaciones de IoT: los registries y los registrars están en una posición única para proporcionar servicios específicos de IoT que contribuyen a la seguridad, privacidad y transparencia de IoT para los usuarios finales. Principalmente, para reducir la probabilidad de secuestro de dominio, que ocurre cuando alguien realiza cambios no autorizados en la información de registro de un dominio, lo que resulta en dispositivos IoT que consultan ese registro modificado y terminan en un servicio diferente (malicioso).

- Además, los conjuntos de datos DNS (consultas DNS) permiten una IoT más transparente para los usuarios finales: Las consultas DNS pueden actuar como una fuente de datos para proporcionar a los usuarios más información sobre los servicios que sus dispositivos IoT usan de manera opaca y que potencialmente procesan sus datos personales. Por lo tanto, es importante proporcionar a los usuarios más información y control sobre los dispositivos de resolución DNS que usan sus dispositivos IoT.

Los autores del reporte SAC105: **The DNS and the Internet of Things: Opportunities, Risks, and Challenges,** consideran que pueden existir tres formas en que IoT puede causar estrés en el DNS:

- Por un lado, la programación hostil para DNS a escala IoT. Una posible causa de una carga adicional en el DNS son los dispositivos IoT que usan el DNS ingenuamente, lo cual puede causar estrés en los resolutores de DNS locales en esas redes porque llenan sus cachés y se quedan sin memoria, lo que resulta en caídas de paquetes o aumento de la latencia de respuesta. Un evento similar sería una gran cantidad de dispositivos IoT que vuelven a estar en línea después de un corte de energía y todos intentan localizar sus servicios remotos casi simultáneamente. Sin embargo, los efectos reales serán difíciles de estimar de antemano porque dependen de varios factores operativos y específicos del sitio, como las concentraciones de dispositivos IoT en las redes, con qué frecuencia resuelven un nombre de dominio y el TTL (time to live) de los nombres de dominio de los servicios de IoT, es decir el tiempo en que demoran los cambios realizados en la configuración del nombre de dominio (crear una zona nueva, cambiar la IP de una zona o subdominio porque cambiamos de servidor donde está la web) para que se reflejen en nuestro navegador. Otra causa puede ser simples errores de programación en el software del cliente IoT o los desarrolladores de software que hacen suposiciones falsas sobre los nombres de dominio.

- Luego, un mayor tamaño y complejidad de ataques DDoS impulsados por botnets IoT: El segundo riesgo que se identifica son las botnets de IoT que pueden atacar el DNS (y otros tipos de infraestructura de Internet) con grandes ataques DDoS coordinados. Una botnet IoT consiste en un conjunto de dispositivos IoT (por ejemplo, cámaras IP y DVR) que se han visto comprometidos por software malicioso. Los dispositivos en una botnet (referidos individualmente como "bots") son controlados por un maestro de botnet, que puede, por ejemplo, indicarles que envíen simultáneamente tráfico a un objetivo específico (por ejemplo, un operador DNS), y así llevar a cabo un ataque DDoS. Los bots infectan los dispositivos de IoT a través de una red al iniciar sesión en ellos o explotando vulnerabilidades de software. Es mucho más difícil erradicar una botnet de IoT que una “botnet tradicional” debido, entre otras cosas, a la naturaleza de los dispositivos y al tiempo que permanecen conectados a la red.

- Finalmente, un mayor número de resolutores abiertos de DNS son aquellos que se han configurado incorrectamente para aceptar consultas DNS de cualquier cliente en Internet, en lugar de restringir el acceso a los clientes dentro del dominio que están destinados a servir, por ejemplo, una red ISP o una red doméstica: Si bien los resolutores abiertos son un problema de larga data, representan un riesgo adicional para IoT. Esto se debe a que una botnet de varios

bots puede lanzar ataques DDoS directos a los operadores de DNS que pueden provocar interrupciones del servicio a gran escala, que potencialmente serían decenas de veces más altas si se amplificaron a través de un conjunto de resolutores abiertos. Otro riesgo es que los ingenieros de dispositivos de IoT pueden tener experiencia limitada con las redes y con la seguridad del dispositivo/red, y como resultado, introducir vulnerabilidades que permiten a los adversarios instalar un resolutor abierto en sus dispositivos, de forma similar a como los bots infectan otros dispositivos IoT. Del mismo modo, los ingenieros de dispositivos de IoT pueden vincular un resolutor de DNS en su software de IoT (por ejemplo, a través de un módulo de terceros) y luego configurarlo incorrectamente accidentalmente. Si bien estos riesgos también existen con las computadoras portátiles y los teléfonos inteligentes, los dispositivos de resolución abiertos pueden ser más persistentes en los dispositivos de IoT porque los dispositivos de IoT suelen ser más difíciles de actualizar y porque a menudo operan sin supervisión, lo cual sólo desaparecerá cuando los dispositivos se desconecten o lleguen al final de su vida útil.

Se consideran un riesgo porque podrían estresar el DNS al reducir el valor del DNS como un servicio confiable. Si bien estos riesgos no son nuevos, se espera que las características de IoT presentarán nuevos desafíos para los operadores de DNS y otros operadores de infraestructura.

Postscapes, en lugar de tratar de ajustar todos los Protocolos IoT sobre los modelos de arquitectura existentes, dividen los protocolos en las siguientes capas para proporcionar cierto nivel de organización:

- Infraestructura
- Identificación
- Comunicación / Transporte
- Descubrimiento
- Protocolo de Datos
- Administración de Dispositivos
- Semántica
- Marco multi-capas
- Seguridad

Una nota adicional merece el protocolo IPv6: los principales desafíos para los desarrolladores de la IoT son que IPv6 no es interoperable con IPv4 en forma nativa y que la mayor parte del software de bajo costo fácilmente disponible para embeber

dispositivos de la IoT solo implementa IPv4. Sin embargo, muchos expertos creen que IPv6 es la mejor opción de conectividad y permitirá que la IoT realice su potencial.

Christopher Yoo se refiere a las oportunidades y desafíos que implica IoT para la privacidad y la seguridad, sosteniendo que la naturaleza distribuida y la presencia de sensores físicos, necesarios para el funcionamiento de IoT, crean nuevas oportunidades y vulnerabilidades desde el punto de vista de la seguridad y la privacidad, como, por ejemplo:

- Naturaleza de información recopilada, cada vez más, personal y sensible: La naturaleza de la información recopilada, cada vez más de tipo personal y sensible, combinada con otro tipo de datos y análisis predictivos (“big data”) pueden permitir inferir niveles detallados de información personal y comportamientos de usuarios.

- Naturaleza distribuida del almacenamiento y procesamiento de datos: Muchos sistemas IoT almacenan y procesan localmente los datos, en los mismos dispositivos, y no los transmiten a un centro de datos. Esto tiene ventajas y desventajas: no tener toda la información almacenada en un único depósito de datos, puede eliminar el incentivo de los atacantes, aunque el almacenamiento descentralizado aumenta la posibilidad de que algunas ubicaciones no mantienen consistentemente los niveles apropiados de seguridad.

- Sensores como un nuevo vector de ataque: El hecho de que los sistemas IoT incorporan necesariamente sensores que recopilan datos del mundo físico los somete a un vector de ataque completamente nuevo. Además de las amenazas tradicionales que existen en línea, atacar a un vector, implica nuevas afectaciones de seguridad para los usuarios. Es importante recordar que un vector de ataque es la ruta o el camino que utiliza un atacante para tener acceso al activo objetivo de ataque.

- Posible corrupción de los dispositivos IoT: debido a que los dispositivos IoT, al ser parcialmente programables y estar conectados a la red, hace que existan más posibilidades de que los atacantes se puedan apropiar de ellos, provocar su mal funcionamiento, e incluso, que sean objeto de ransomware y extorsionar a los usuarios mediante el secuestro de dichos dispositivos.

Todo ello, debido a que, la mayoría de los sistemas IoT no se diseñaron teniendo en cuenta la seguridad. Por lo tanto, el diseño adecuado de privacidad y seguridad de IoT debe concebirse como un balance que intente alcanzar el equilibrio adecuado entre funcionalidad y seguridad.

También será necesario determinar en los próximos años, cuáles serán las reglas o regulaciones para determinar quién asumirá la responsabilidad por los daños causados por los dispositivos de IoT.

Internet Society considera que con relación a IoT es necesario contar con estándares apropiados, modelos de referencia y mejores prácticas que ayuden a frenar la proliferación de dispositivos que podrían alterar a Internet. El uso de estándares genéricos, abiertos y ampliamente disponibles (como el Protocolo de Internet) –en contraposición a implementaciones técnicas propietarias de IoT- como componentes de los dispositivos y servicios de la IoT permitirá mayores ventajas para los usuarios, más innovación y más oportunidades económicas.

Lo anterior va de la mano con productos o servicios de IoT que sean interoperables, que tengan flexibilidad en su integración y que no represente gran complejidad en cuanto a su propiedad y preocupación con respecto a posibles dificultades para cambiar de proveedores de tecnología.

Además, los dispositivos de la IoT mal diseñados y configurados pueden tener consecuencias negativas para los recursos de red a los cuales se conectan y para Internet en un sentido más amplio.

En cuanto a la regulación de IoT, especialistas defienden la necesidad de una normativa flexible, basada en principios, guías y directrices, ya que se considera que es la más eficaz y fácil de consensuar a nivel internacional, con la participación de todos los interesados en su confección. Otorgan un papel predominante a los principios de buena fe, que obliga a proporcionar información adecuada y suficiente a los usuarios, de privacidad, que exige recabar el consentimiento de los participantes, de proporcionalidad, de seguridad y de neutralidad tanto respecto de la red como de la tecnología. (Susana Navas Navarro, “El Internet de las cosas”).

Algunos autores consideran que no sólo se deben elaborar legislaciones que regulen IoT sino que principalmente, se deben establecer mecanismos de gobernanza con la participación de los múltiples actores, que permitan el desarrollo de políticas públicas sobre IoT e incentivos económicos para prevenir problemas en su implementación, antes que los mismos ocurran.

Ejemplo de regulaciones podría ser el siguiente:

- En 2020 entrará en vigor la California IoT Cybersecurity Law que requiere que los fabricantes de dispositivos conectados a Internet, como televisores, teléfonos,

juguetes, electrodomésticos y enrutadores, se aseguren de que sus productos tengan "características de seguridad razonables". Estas características de seguridad deberían poder proteger la información confidencial del cliente del acceso no autorizado.

"La falta de funciones de seguridad básicas en los dispositivos conectados a Internet socava la privacidad y la seguridad de los consumidores de California y permite a los piratas informáticos recurrir a la electrónica de consumo cotidiana contra nosotros", dijo el senador Jackson, promotor de esta iniciativa.

Módulo 6.2: Blockchain

El blockchain ("cadena de bloques") es una base de datos encriptada, descentralizada y compartida, que sirve de repositorio de información público, irreversible e incorruptible según Wright y De Filippi.

Características de la tecnología:

- **Descentralizada:** porque no es necesario confiar su funcionamiento a una autoridad central. Es la propia red interconectada de computadoras, interoperables, las que garantizan que una transacción pueda realizarse eficazmente. Todas las computadoras o nodos almacenan la información de todas las cadenas de transacciones y validan las nuevas.
- **Abierta:** porque cualquier persona puede acceder a ella.
- **Pública:** porque es posible ver y rastrear todas las relaciones que se producen.
- **Segura:** porque se considera que sólo podría caer si un ataque logra capturar más del 50% de las computadoras que conforman el sistema (nodos), lo cual, para sus defensores, se considera prácticamente imposible.
- **Ofrece infinitas posibilidades de negocios.**

Los términos algoritmo y protocolo a menudo se emplean de manera indistinta, sin embargo, no son la misma cosa. En términos simples, podemos definir un protocolo como las reglas primarias de una blockchain; y el algoritmo, como el mecanismo a través del cual dichas reglas serán seguidas.

Una red blockchain siempre se erigirá sobre un protocolo que definirá la supuesta forma de funcionamiento del sistema. Por ese motivo, todos los elementos de éste, así como los participantes de la red, deberán respetar las reglas del protocolo subyacente.

Mientras el protocolo determina cuáles son las reglas, el algoritmo le dice al sistema qué pasos seguir para cumplir con las mismas y producir los resultados deseados. Por ejemplo, el algoritmo de consenso de una blockchain es lo que determina la validez de las transacciones y bloques. Así, Bitcoin y Ethereum son protocolos, mientras que Proof of Work y Proof of Stake son sus algoritmos de consenso.

Respecto a Proof of Work: se asegura de que los mineros sólo sean capaces de validar un nuevo bloque de transacciones y añadirlo al blockchain, si los nodos distribuidos de la red alcanzan consenso y aceptan el hash block provisto por el minero como una prueba de trabajo válida.

Luego, Proof of Stake reemplaza el minado PoW por un mecanismo en que los bloques son validados de acuerdo con el "stake" (o sea cantidad de monedas acumuladas) de los participantes.

Los principios del blockchain son:

- **Integridad en la red:** la confianza es intrínseca. La integridad está cifrada en todas y cada una de las etapas del proceso y distribuida, y no depende de cada miembro individualmente.

- **Poder distribuido:** el sistema distribuye poder por una red de iguales sin que haya ningún punto de control. Las partes no pueden apagar el sistema por sí solas. Si una autoridad central logra inhabilitar o expulsar a un individuo o a un grupo, el sistema sobreviviría. Si la mitad de la red intentara dominar al conjunto, todo el mundo lo vería.

- **El valor como incentivo:** el sistema hace coincidir los incentivos de todos los participantes.

- **Seguridad:** las medidas de seguridad están integradas en la red sin puntos flacos y no sólo garantizan la confidencialidad sino también la autenticidad de todas las actividades y la imposibilidad de que nos sean denegadas. Todo el que quiera participar debe usar criptografía y las consecuencias de comportarse mal sólo las sufre la persona que se comporta mal.

- **Privacidad:** la gente debería poder controlar sus datos, tener el derecho de decidir sobre su identidad, y cuándo, cómo y cuánta compartir con los demás. Respetar el derecho a la privacidad no es lo mismo que respetar la privacidad. Son necesarias las dos cosas.

- **Derechos preservados:** los derechos de propiedad son transparentes y legítimos. Las libertades individuales están reconocidas y son respetadas.

- **Inclusión:** la economía funciona mejor cuando funciona para todos. Eso significa eliminar obstáculos que dificulten la participación y crear nuevas plataformas que hagan posible un capitalismo distribuido.

Dentro de las aplicaciones de blockchain, se pueden mencionar a las criptomonedas, de las cuales el Bitcoin es el más conocido.

El bitcoin es una moneda electrónica, virtual, descentralizada, finita y voluble, que utiliza la técnica de encriptación para su seguridad. Es una moneda virtual, aunque no tiene las consideraciones del dinero ya que no es ampliamente aceptada como medio de pago de bienes y servicios; no tiene un valor intrínseco, ya que el que pueda tener deriva exclusivamente de la creencia de los usuarios de que podrán cambiarla por bienes, servicios o por divisas nacionales. Asimismo, por su naturaleza esencialmente transnacional y la ausencia de un organismo acuñador, dificultan su regulación por parte de las autoridades nacionales y suponen un reto para los bancos centrales.

Algunas de las aplicaciones al blockchain son:

- Contratos inteligentes (“Smart Contract”): programa informático, basado en la tecnología de cadena de bloques, que se auto ejecuta sin la intervención de las partes.
- Blockchain aplicado al terreno de la propiedad, para lo que se requeriría la digitalización de los bienes y derechos. Se incluye el Blockchain en registros públicos (de la propiedad, mercantil y de la propiedad intelectual).

- Blockchain aplicado a la economía colaborativa para terminar con la intermediación.

Los riesgos de regular prematuramente –antes de entender bien las implicaciones- pueden tener profundas consecuencias.

Regulación no es lo mismo que gobernanza. Como se vio en el módulo 1, la regulación consiste en el establecimiento de normas, reglas o leyes dentro de un determinado ámbito. El objetivo de la regulación es mantener un orden, llevar un control y garantizar los derechos de todos los integrantes de una comunidad. Las

regulaciones de Internet pueden estar incluidas tanto en tratados internacionales y en leyes locales o nacionales como en los términos y condiciones de uso de los servicios de Internet, plataformas, aplicaciones, estándares y protocolos técnicos, entre otros.

Mientras que la gobernanza, en sentido genérico, hace referencia a la suma de las múltiples maneras como los individuos y las instituciones, públicas y privadas, manejan sus asuntos comunes. Es un proceso continuo mediante el que pueden acomodarse intereses diversos y conflictivos y adoptarse alguna acción cooperativa. Incluye las instituciones formales y los regímenes con poder para imponer obediencia, así como arreglos informales que las personas y las instituciones han acordado.

Se propone la creación de Redes de gobernanza de blockchain, la cual debe ser inclusiva y abierta a la participación de todos los grupos de actores interesados relevantes; meritocrática, transparente, poniendo a disposición del público toda la información, y las decisiones deberían tomarse, en la medida de lo posible, por consenso para que el resultado sea lo más legítimo posible.

Módulo 6.3: Inteligencia Artificial y los desafíos de gobernanza que presenta.

El desarrollo de los algoritmos se encuentra en la base del desarrollo de la computación y de los propios protocolos de Internet. Es incluso muy relevante contextualizar que los primeros experimentos en torno a la futura Internet con la conmutación de paquetes, como el concepto de Inteligencia Artificial fueron apoyados por la misma agencia de investigación en sus orígenes en la década de 1950 – DARPA (Defense Advanced Research Projects Agency).

Hay dos acepciones fundamentales de un algoritmo.

- Desde su caracterización formal es considerado un constructo matemático a los efectos de lograr un propósito bajo ciertos suministros.
- Desde su empleo en el discurso público alude a que debe ser entonces implementado y ejecutado para llevar a cabo una acción y obtener ciertos efectos.

La **IA** es el estudio de los agentes que reciben información del entorno y desarrollan acciones.

La Inteligencia se operacionaliza como la creación de agentes racionales, que significa que el agente está “haciendo lo correcto” a partir de una métrica determinada.

El **machine learning** es caracterizado por la capacidad de definir o modificar reglas de toma de decisión de manera autónoma. Estos mecanismos parten de una serie más o menos grande, más o menos estructurada de datos (las llamadas “experiencias pasadas”) y a partir de estos datos descubren patrones preexistentes en ellos, lo que les permite predecir qué va a ocurrir ante la aparición de un nuevo dato, o realizar tareas varias en función de la experiencia pasada acumulada, del aprendizaje realizado.

El proceso de aprendizaje varía, puede ser guiado o automático (es decir puede haber humanos que corrijan el avance en el aprendizaje -maestros- o puede aprender la máquina interactuando consigo misma), activo o pasivo (esto es puede darse en medio de una constante interacción con el entorno que retroalimenta el proceso de aprendizaje, o puede ser puramente observacional, esto es que la máquina adquiere información antes de empezar a actuar), puede ser orientado a un fin (que la máquina aprenda a realizar una tarea específica) o simplemente que encuentre patrones inesperados entre los datos preexistentes (la llamada minería de datos).

El aprendizaje puede ser supervisado (vía *inputs* etiquetados a mano) o no supervisado (o sea que el algoritmo mismo define los modelos que mejor encajan para dotar de sentido al conjunto de *inputs*). En ambos casos, el algoritmo define las reglas de toma de decisión para manejar nuevos *inputs*. En consecuencia, el operador humano no necesita comprender la razón subyacente a las reglas de toma de decisión producidas por el algoritmo. Como contrapartida de la opacidad en cuestión, la transparencia es generalmente deseada, dado que los algoritmos que son poco predecibles o explicables son difíciles de controlar, monitorear o corregir.

¿Por qué los avances de la Inteligencia Artificial constituyen un tema de gobernanza? Para responder a esta pregunta debemos empezar por identificar los principales problemas que se derivan de las decisiones automatizadas.

El primer elemento que surge como problemático es el sesgo. Como fue mencionado, sobre la base del ciclo de vida del aprendizaje automático, a saber: datos de entrenamiento, entrenamiento del algoritmo, implementación en situaciones reales, evaluación de resultados y ajuste de parámetros, hay una serie de riesgos asociados que condicionan este sesgo. Con sesgo hacemos referencia al sesgo algorítmico que ocurre cuando un sistema informático refleja los valores implícitos de los humanos que lo crearon. Estos son:

Según Friedman y Nissenbaum ,Bias in Computer Systems 1996, los valores sociales pre-existentes, de las instituciones sociales, prácticas y actitudes de las que surge la

tecnología, las restricciones técnicas que surgen de limitaciones tecnológicas o decisiones de diseño, que favorecen grupos particulares sin valores directrices subyacentes. Por ejemplo, cuando un listado alfabético de cadenas hoteleras conduce a incrementar las ventas de aquellas que están antes en el alfabeto; por último, encontramos los aspectos emergentes de un contexto de uso que están ligados a avances en el conocimiento o cambios deliberados en los usos del sistema, por ejemplo, los sistemas de apoyo a decisiones clínicas están inevitablemente sesgados hacia tratamientos incluidos en su arquitectura de decisión.

La privacidad es otra dimensión que se ve afectada por el uso extensivo de la IA. El documento para una IA confiable de la Unión Europea recomienda proteger los datos personales en todas las instancias del desarrollo y uso de sistemas IA.

Para lograr esto, el primer paso en cualquier diseño debería ser anonimizar los datos personales, de tal manera de que no sean identificables con la persona de carne y hueso que los introdujo en el sistema a los efectos de usar esos datos para generar sistemas de aprendizaje automatizado con ellos. De esta manera los datos mismos serían útiles para generar sistemas de IA, pero no correría ningún peligro ni estaría en desventaja la persona que cedió sus datos. Por ejemplo, se podría generar un sistema que ayude a las personas diabéticas a regular sus dosis diarias de insulina, sin que sea identificable qué personas particulares diabéticas o no cedieron sus datos para crear la app en cuestión, esto es, sobre la base de los datos de quiénes se hizo correr el algoritmo de aprendizaje para generar el modelo.

Es una exigencia ética de la tecnología de la inteligencia artificial comunicar y difundir sus características y aplicaciones para que los ciudadanos puedan tomar decisiones informadas. La transparencia o comprensión de los sistemas de IA está directamente asociada a la autonomía de la vida humana. Si no hay una comprensión mínima de los sistemas tecnológicos que incorporan IA, no se podrá garantizar el ejercicio de la autonomía, puesto que los ciudadanos deberán ceder siempre su autoridad a los expertos y no podrán juzgar lúcidamente por sí mismos respecto de la adopción o no de esa clase de tecnologías. Sin comprensión de la tecnología no hay evaluación juiciosa y adecuada a la realidad de la misma.

Los componentes primarios de la transparencia son: la accesibilidad y) la comprensibilidad de la información (según Glenn y Monteith 2014; y Kitchin 2016).
En esta línea, es importante remarcar que la transparencia podría ser una demanda acerca del **algoritmo** o acerca del **uso** del sistema en cuestión.

Considerando las limitaciones de la “caja negra”, los desarrolladores podrían cuestionar la plausibilidad de ser transparentes con los algoritmos de aprendizaje automático. En ocasiones, sostienen que ellos mismos no podrían descifrar o hacer ingeniería reversa acerca de cómo operan algoritmos específicos después de haberlos entrenado con vastos conjuntos de datos. El pretendido valor de estos algoritmos reposa en su habilidad para resolver la toma de decisión de un modo que sería superior a –en tanto optimizaría- las posibilidades de procesamiento humano. La transparencia algorítmica no contribuiría genuinamente, por tanto, a la comprensión de los usuarios.

Con relación a la transparencia en el uso, se trata de hacer comprensible para el usuario lo suficiente (aunque no todo) de las decisiones tomadas al momento del diseño del sistema, de tal manera de que el usuario pueda desarrollar una confianza en su fiabilidad y justicia subyacente. En general, esto supone dar una cierta información acerca de cómo se desarrolló el sistema, sin tecnicismos, y acorde a la cultura técnica del usuario. Dado que el grado y tipo de información que los distintos usuarios pueden asimilar es variable en diversos grupos humanos, la transparencia es relativa, varía entre grupos y constantemente se requiere de expertos, tanto en diseño informático como científicos sociales que permitan realizar la tarea de “traducción” de la especificación técnica del sistema para los usuarios efectivos.

La fundación **World Wide Web** distingue la responsabilidad algorítmica como la obligación de los diseñadores de algoritmos de dejar en claro cuáles son los daños potenciales de sus desarrollos- de la justicia algorítmica, vinculada con la capacidad para reparar daños ocasionados. Esta diferenciación ayuda a echar luz sobre la necesidad de pensar acerca de la responsabilidad antes de que ocurra cualquier daño y cómo aquellos que piensan, diseñan y desarrollan los algoritmos deben prever también las consecuencias indeseables de sus trabajos.

Es fundamental problematizar entonces la dimensión de la gobernanza de la IA. ¿Qué constituye la gobernanza de la IA?

Una definición posible son las reglas intersubjetivas que definen, constriñen y moldean las expectativas en torno a las propiedades fundamentales de un *agente artificial* (en base a Gahnberg, 2019).

¿Por qué es importante establecer mecanismos para la gobernanza de la IA?

- se le pueden delegar decisiones;
- moldean nuestro entendimiento del mundo;

- moldean nuestro comportamiento

De forma más inquietante, el **Machine Learning** lo puede hacer de forma impredecible y opaca. Esto constituye un factor fundamental para abordar su gobernanza de forma de establecer mecanismos que permitan un desarrollo de esta tecnología, pero persiguiendo el objetivo del desarrollo sostenible, y respetuoso de los derechos humanos.