Cooperación Española
CONOCIMIENTO / INTERCOONECTA

Curso de Especialización Judicial
"Ciberdelincuencia: tratamiento preventivo, procesal y sustantivo desde una perspectiva internacional". 4ª Edición

3.2.- LA PRUEBA DIGITAL. IMPACTO DE LA INTELIGENCIA ARTIFICIAL

Autor
Joaquín Delgado Martín
Magistrado Sala Penal de la Audiencia Nacional
ESPAÑA

Consejo General del Poder Judicial
Escuela Judicial

Aula Iberoamericana
Consejo General del Poder Judicial

ESQUEMA GENERAL DE LA INTERVENCIÓN:

Previo.- ¿De qué estamos hablando?: fuentes de la prueba digital

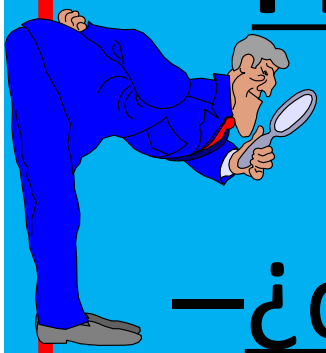
1.Interceptación de comunicaciones

2.Registros digitales

3. Fuentes abiertas

4. Datos en poder de proveedores de servicios

PREVIO.- E-EVIDENCE O PRUEBA DIGITAL



- ¿de qué estamos hablando?: Toda información producida, almacenada o transmitida por medios electrónicos (en formato digital), que pueda tener efectos para acreditar hechos en el proceso

FUENTES DE DATOS PARA LA PRUEBA DIGITAL

Publicación en Web

Prestadores de servicios

Dispositivo 1

Autor del delito/víctima

Dispositivo 2

Víctima/autor del delito

FUENTES DE DATOS PARA LA PRUEBA DIGITAL

Prestadores de servicios:
- **Obligación de conservación**
- Conservación por propia iniciativa
- **Orden de "congelación"**

Datos en Web:
- **Fuentes abiertas**
- Proceso de comunicación

Proceso comunicación (acceso datos en transmisión: **interceptación de comunicaciones**

Dispositivo 1:
- Aprehensión física:
  - Contenido del dispositivo: **Registro de dispositivos**
  - Contenido accesible desde el dispositivo: **Registro de información accesible**
- No aprehensión física:
  - **Registro remoto** (troyanos..)

Dispositivo 2

MEDIDAS RESTRICTIVAS DDFF: PRINCIPIOS

- — **Principio de especialidad**: *«exige que la medida esté relacionada con la **investigación de un delito concreto**. No podrán autorizarse medidas de investigación tecnológica que tengan por objeto prevenir o descubrir delitos o despejar sospechas sin base objetiva»* [art. 588 bis a) 2 LECRIM]. De esta manera, las medidas de investigación tecnológica prospectivas sobre la conducta de una persona o grupo están prohibidas.
- **— Principio de idoneidad:** *«servirá para definir el ámbito objetivo y subjetivo y la duración de la medida en virtud de su utilidad»* [art. 588 bis a) 3 LECRIM].
- **— Principios de excepcionalidad y necesidad:** de conformidad con el art. 588 bis a) 4 LECRIM, *«solo podrá acordarse la medida: a) cuando no estén a disposición de la investigación, en atención a sus características, otras medidas menos gravosas para los derechos fundamentales del investigado o encausado e igualmente útiles para el esclarecimiento del hecho, o b) cuando el descubrimiento o la comprobación del hecho investigado, la determinación de su autor o autores, la averiguación de su paradero, o la localización de los efectos del delito se vea gravemente dificultada sin el recurso a esta medida»*.
- **— Principio de proporcionalidad:** *«las medidas de investigación reguladas en este capítulo solo se reputarán proporcionadas cuando, tomadas en consideración todas las circunstancias del caso, el sacrificio de los derechos e intereses afectados no sea superior al beneficio que de su adopción resulte para el interés público y de terceros. Para la ponderación de los intereses en conflicto, la valoración del interés público se basará en la gravedad del hecho, su trascendencia social o el ámbito tecnológico de producción, la intensidad de los indicios existentes*

1.- INTERCEPTACIÓN DE COMUNICACIONES ELECTRÓNICAS

INTERCEPTACIÓN

- CONCEPTO: captación **en tiempo real** del **contenido** y **datos** asociados de una comunicación, tanto de telefonía (fija o móvil) como de cualquier tipo de red de datos, sin interrumpir el curso de la misma, para la obtención de datos útiles para la investigación y prueba del delito .
- En estos casos resulta afectado el derecho fundamental al secreto de comunicaciones
- Régimen jurídico español se contiene en los arts. 588 ter a y ss LECRIM

Interceptación: ámbito subjetivo

A. Sujeto activo

- Principio general: el **Juez**.
- Reserva jurisdiccional

B. Sujetos pasivos

- **Del investigado**habitual u ocasionalmente utilizados por el investigado.
- **De tercero**: los pertenecientes a una tercera persona, siempre que concurran alguno de los siguientes supuestos: 1.º que exista constancia de que el sujeto investigado se sirve de aquella para transmitir o recibir información, o 2.º el titular colabore con la persona investigada en sus fines ilícitos o se beneficie de su actividad.
- **De la víctima:** también podrán intervenirse los terminales o medios de comunicación de la víctima cuando sea previsible un grave riesgo para su vida o integridad.

C. Utilización maliciosa por terceros. Routers. Ordenadores zombies-botnets

- También es posible cuando el dispositivo objeto de investigación sea utilizado maliciosamente por terceros por vía telemática, sin conocimiento de su titular

Interceptación: ámbito objetivo

A. ¿Qué delitos pueden ser investigados?

- Gravedad en torno a la pena, el tipo de delito y/o otras circunstancias (por ejemplo organización criminal)

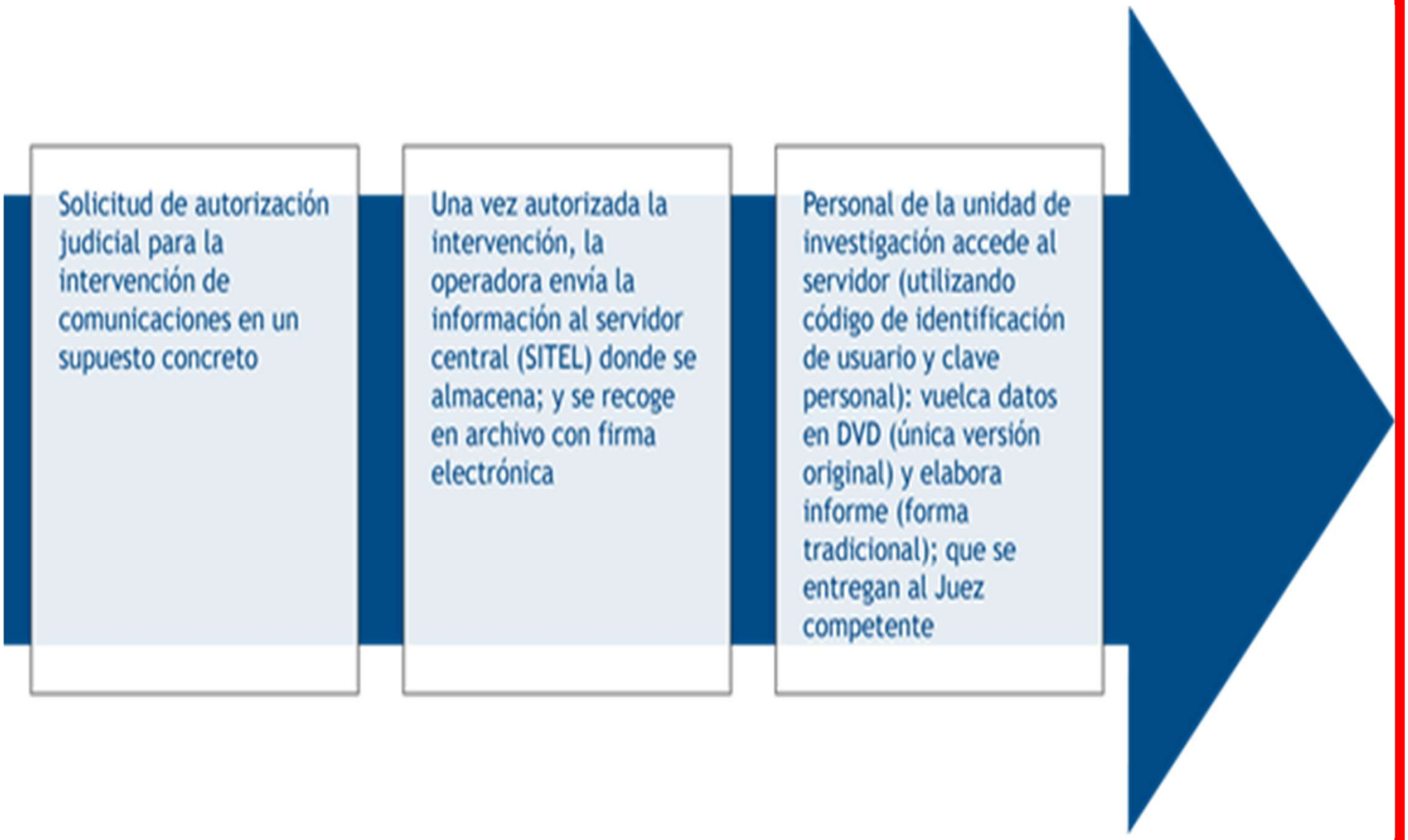
B. ¿Qué dispositivos pueden ser intervenidos?

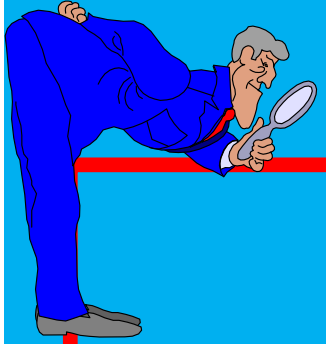
- Los terminales o sistemas de comunicación

C. ¿A qué información puede accederse?

- La extensión de una concreta intervención ha de ser establecida por la resolución judicial autorizante, como medio para garantizar los principios de idoneidad, excepcionalidad y necesidad. Puede abarcar alguna de las siguientes informaciones:
  - El **contenido de las comunicaciones** en las que participe el sujeto investigado, ya sea como emisor o como receptor.
  - Los **datos electrónicos de tráfico** o asociados al proceso de comunicación: todos aquellos que se generan como consecuencia de la conducción de la comunicación a través de una red de comunicaciones electrónicas, de su puesta a disposición del usuario, así como de la prestación de un servicio de la sociedad de la información o comunicación telemática de naturaleza análoga
  - Los **datos que se produzcan con independencia** del establecimiento o no de una concreta comunicación.

Procedimiento de interceptación: ejecución SITEL





2.- REGISTROS:

- De dispositivos electrónicos
- Datos en la nube
- Registros remotos

2.1.- REGISTRO DE DISPOSITIVOS ELECTRÓNICOS

¿QUÉ SON LOS DISPOSITIVOS ELECTRÓNICOS?

¿QUÉ SON DISPOSITIVOS ELECTRÓNICOS?

- ***Lenguaje binario***. 00011010001001000100110001
  - Necesario un dispositivo para convertirlo en lenguaje alfabético o en imágenes, videos, tablas….
- **Gran variedad**:
  - Aparatos de tecnología digital (teléfonos móviles, smartphones, tabletas, ordenadores, GPS…)
  - Medios de almacenamiento masivo de memoria (dispositivos USB, ZIP, DVD…)
- **Antes en otros formatos, ahora en digital**:
  - fotografías, videos
- **Número creciente**:
  - «Internet de las cosas» (Iot-Internet of Things)
  - 5-G

¿EN QUÉ CONSISTE Y PARA QUÉ VALE EL REGISTRO DE DISPOSITIVOS?

DATOS EN DISPOSITIVOS ELECTRÓNICOS

- ***ACCESO A LA INFORMACIÓN***.
  - informaciones muy heterogéneas: registros de actividad (logs), bases de datos, comunicaciones digitales (e-mails, SMS, comunicaciones de mensajería instantánea,…), imágenes (formato jpg, png,…), documentos ofimáticos (Office, Adobe,…),
  - Pueden resultar de gran importancia para la investigación y ulterior prueba de hechos relevantes en un proceso judicial (penal, civil, laboral, mercantil, familia…..)
- **GRAVEDAD DE LA INTROMISIÓN**:
  - Por el gran número de datos contenidos en un dispositivo, que además afectan a largos periodos de tiempo
  - Por los datos aisladamente considerados; muchos datos pueden afectar al núcleo más sensible de la intimidad
  - Por la información generada por los datos agregados analizados en su conjunto (perfiles….)
  - Algunos datos pueden afectar al secreto de comunicaciones

¿QUÉ DERECHOS FUNDAMENTALES SON AFECTADOS?

A. Derechos fundamentales afectados

I. Intimidad personal

II. Secreto de las comunicaciones

III. Derecho a la autodeterminación informativa en el ámbito de la protección de datos personales

B. Protección eficaz del entorno virtual del afectado

A. Derecho constitucional de nueva generación

- "Derecho fundamental a la garantía de confidencialidad e integridad de los grupos informáticos" (STC alemán 27-2-2008): "*protege la vida privada y personal de los sujetos de los derechos fundamentales contra el acceso por parte del Estado en el ámbito de las tecnologías de la información, en la medida en que el Estado posea acceso al sistema de tecnologías de la información en su conjunto y no sólo a los acontecimientos individuales o a los datos almacenados*".
- "Derecho al propio entorno virtual" (STS 342/2013)

B. Tratamiento unitario en la Ley española (reforma 2015): artículos 588 sexies a-e LECRIM

¿CUÁNDO ES LÍCITO EL REGISTRO DEL DISPOSITIVO?

ACCESO LÍCITO A DISPOSITIVO

1.- Autorización y control judicial

- Principios de especialidad, idoneidad, excepcionalidad y necesidad y proporcionalidad
- Motivación de la resolución judicial
- Investigación de cualquier delito: filtro de la proporcionalidad
- Control judicial

2.- Intervención policial en casos de urgencia

- Urgencia y necesidad
- Proporcionalidad

3.- Consentimiento del afectado

URGENCIA POLICIAL: PRESUPUESTOS

- La **urgencia** en el acceso a los datos. La STC 115/2013, de 9 de mayo, entendió que la intervención resultó urgente por la necesidad de averiguar la identidad de alguna de las personas que huyeron al ser sorprendidas *in fraganti* custodiando un alijo de droga, para proceder a su detención de tal manera que no se sustrajeran definitivamente a la acción de la justicia.

- La **necesidad** de obtener la información. El registro ha de resultar estrictamente necesario para la finalidad de la investigación, esto es, solamente podrá acordarse cuando el mismo fin no pueda lograrse por otro medio menos gravoso para el afectado. En definitiva, nos encontramos ante una cláusula de subsidiariedad, de tal manera que el medio seleccionado para alcanzar el fin no pueda ser suplido por otro igualmente eficaz, pero que no restrinja el derecho fundamental o lo haga de una manera menos gravosa.

- La **proporcionalidad** en la actuación. Como afirma la STC 115/2013, "*se trató de una medida ponderada o equilibrada, por derivarse de ella más beneficios o ventajas para el interés general que perjuicios sobre otros bienes o valores en conflicto, dada la naturaleza y gravedad del delito investigado y la leve injerencia que comporta en el derecho a la intimidad del recurrente el examen de la agenda de contactos de su teléfono móvil*".

Consentimiento del afectado

- Puede legitimar la injerencia
- Aunque puede ser revocado
- Expreso o tácito (actos concluyentes)
- Libre e informado (problema de la presión/atmósfera policial)

¿CUÁLES SON LOS EFECTOS DE UN REGISTRO ILÍCITO?

NULIDAD DE PLENO DERECHO:

- No despliega **ningún efecto en el proceso**. La sentencia no podrá fundamentarse en ninguno de los siguientes elementos:
  - Datos que pudieran resultar del acceso ilícito
  - Percepciones de los sujetos que hubieran intervenido en la actuación de acceso ilícito (por ejemplo, una entrada y registro en domicilio con acceso al contenido de los ordenadores encontrados en el proceso penal); aunque los mismos declaren en juicio como testigos.
  - Resultados de las pruebas que se proyecten sobre los datos encontrados mediante el acceso ilícito: dictámenes periciales, o las declaraciones testificales que tengan por objeto el reconocimiento de los mencionados datos.
- **No puede ser subsanada** ni ser objeto de convalidación de ninguna manera.
- **No debe ser incorporada al proceso**; y si se ha tenido entrada en él, el ordenamiento contempla la posibilidad de que sea excluida del mismo o no ser valorada por el tribunal de enjuiciamiento.

CADENA DE CUSTODIA Y PRESERVACIÓN DE LOS DATOS

CADENA DE CUSTODIA EN REGISTRO DE DISPOSITIVOS

- **Concepto de cadena de custodia:**
  - es el procedimiento, oportunamente documentado, que permite constatar la identidad, integridad y autenticidad de los vestigios o indicios de un hecho relevante para el asunto, desde que son encontrados hasta que se aportan al proceso como pruebas
  - Datos de dispositivos: **autenticidad** e **integridad**

PRESERVACIÓN DATOS

- **Autenticidad**: coincidencia de su autor aparente con su autor real (tradicional). En el ámbito de la prueba electrónica: propiedad o característica consistente en que se garantiza la autenticidad del origen de los datos, es decir, se garantiza la fuente de la que proceden los datos
- **Integridad**: propiedad o característica consistente en que los datos (activo de información) no han sido alterados de manera no autorizada

VOLCADO o CLONADO

- Volcado o clonado
  1. Realización de una copia espejo o bit a bit de la información original
     - Es copia física y no copia lógica
     - “Original”+”copia 1ª”+ ”copia 2ª”
     - mediante una herramienta de hardware, de tal forma que se lleva a cabo una copia física del contenido
  2. Fijación del código hash que se calcula a partir de un algoritmo de cifrado estándar que posibilita concluir que los datos hallados en el dispositivo en el momento de su aprehensión no han sido objeto de ulterior manipulación

VOLCADO DE DATOS

- **Lugar:**
  - In situ (allí donde está el soporte); aunque no necesariamente
- **Tiempo:**
  - Durante el registro o posteriormente
- **Garantías**
  - Técnicas: utilización de instrumentos tecnológicos y procedimientos adecuados (estandarización/homologación)
  - Jurídicas: presencia de testigos y/o fedatario público

2.2.- ACCESO A DATOS EN LA NUBE: REGISTRO AMPLIDO

(www.incibe.es)

DATOS ALMACENADOS EN NUBE

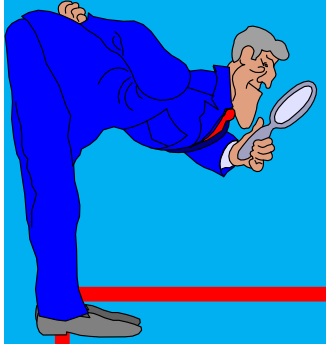
- SUPUESTOS:
  1. Dispositivo abierto o el acceso a su contenido es posible sin uso de claves o contraseñas: SI REGISTRO INFORMACIÓN ACCESIBLE
  2. La autoridad pública conoce las claves de forma legítima (suministrada por titular, análisis forense del dispositivo…): SI REGISTRO INFORMACIÓN ACCESIBLE
  3. Dispositivo cerrado y no se conocen las claves: NO REGISTRO INFORMACIÓN ACCESIBLE (registro remoto)
- PROBLEMA: LOCALIZACIÓN DATOS FUERA TERRITORIO
  - Admitido por normativa española
- RÉGIMEN JURÍDICO REGISTRO INFORMACIÓN ACCESIBLE
  - Autorización judicial inicial: que abarque la información accesible
  - Autorización judicial sobrevenida
  - Intervención policial de urgencia + control judicial posterior

2.3.- REGISTROS REMOTOS

- **Concepto**: la utilización de datos de identificación y códigos, así como la instalación de un software, que permitan, de forma remota y telemática, el examen a distancia y sin conocimiento de su titular o usuario del contenido de un ordenador, dispositivo electrónico, sistema informático, instrumento de almacenamiento masivo de datos informáticos o base de datos
- **Modalidades**:
  1. Mediante el uso de **códigos u otros elementos identificativos**, pero sin la instalación en el mismo de software alguno
  2. Mediante la previa instalación en el sistema investigado de un **software** (los denominados «programas troyanos») que permite a las autoridades escanear un disco duro y demás unidades de almacenamiento y remitir de forma remota y automatizada el contenido del mismo al informático de la autoridad responsable de la investigación

INJERENCIA GRAVE: RÉGIMEN ESTRICTO

- Solamente es posible en la investigación de alguno de los siguientes delitos (**listado exhaustivo de infracciones penales**):
  - a) Delitos cometidos en el seno de organizaciones criminales.
  - b) Delitos de terrorismo.
  - c) Delitos cometidos contra menores o personas con capacidad modificada judicialmente.
  - d) Delitos contra la Constitución, de traición y relativos a la defensa nacional.
  - e) Delitos cometidos a través de instrumentos informáticos o de cualquier otra tecnología de la información o la telecomunicación o servicio de comunicación
- Reserva jurisdiccional
- Duración. De conformidad con el art. 588 septies c), la medida tendrá una duración máxima de un mes, prorrogable por iguales períodos hasta un máximo de 3 meses



3.- DATOS CONSERVADOS POR LOS PROVEEDORES DE SERVICIOS

Proveedores de servicios (concepción amplia=Convenio Budapest)

1. **Servicios de comunicaciones electrónicas**
   a. **Servicio de acceso a internet**
      - Los Internet Service Providers (ISP) son aquellas entidades que brindan conexiones y servicios de Internet a individuos y organizaciones;
      - Además de proporcionar acceso a Internet, los ISP también pueden proporcionar paquetes de software (como navegadores), cuentas de correo electrónico y un sitio web personal o página de inicio

   b. **Servicio de comunicaciones electrónicas interpersonales**
      - Tradicionales (por ejemplo, telefonía de voz, SMS, servicio de acceso a Internet);
      - Nuevos servicios basados en Internet que permiten comunicaciones interpersonales como voz sobre IP, mensajería instantánea y correo electrónico basado en servicios web

2. **Servicios de la sociedad de la información**
   - Gran variedad de proveedores de servicios conocidos como redes sociales (por ejemplo, Facebook y Twitter), servicios en la nube (por ejemplo, Microsoft, Dropbox o Amazon Web Services), mercados online (por ejemplo, eBay o Amazon) u otros proveedores de servicios de alojamiento de datos (por ejemplo, Bluehost).

3. **Servicios de infraestructura de internet**
   - De asignación de nombres de dominio de internet y de direcciones IP

CLASES DE DATOS

- Datos de naturaleza **dinámica**: son los que se generan durante un proceso de comunicación
- Datos de naturaleza **estática**: se almacenan por los prestadores de servicios para posibilitar esas comunicaciones, pero no se generan como consecuencia de una comunicación concreta

- Acceso a datos **junto con interceptación del contenido de la comunicación**
  - Se debe individualizar tanto la solicitud como la decisión judicial en relación con los datos
  - Régimen jurídico de la interceptación del contenido (artículos 588 ter)
- Acceso cuando la comunicación ya ha terminado o porque no ha llegado a existir (llamadas frustradas)

CLASES DE DATOS: DIFERENTE AFECTACIÓN DDFF

DISTINCIÓN TRADICIONAL

- Datos de suscripción
- Datos de tráfico
- Datos de contenido

REGLAMENTO E-EVIDENCE

- Datos de suscripción
- Datos de acceso

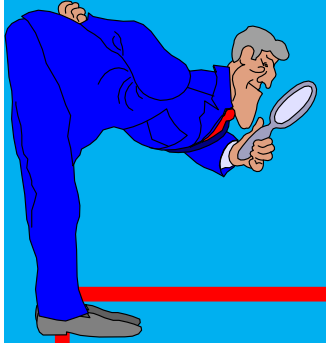
.....

- Datos transaccionales
- Datos de contenido

DISTINCIÓN RELEVANTE PARA EL JUICIO DE PROPORCIONALIDAD

- **Datos de los abonados**: cualquier dato en relación con:
  - (a)la identidad del abonado o cliente, como nombre, fecha de nacimiento, dirección postal o geográfica, facturación y pagos, teléfono o dirección de correo electrónico;
  - (b)el tipo de servicio y su duración, incluidos los datos técnicos que identifiquen las medidas técnicas correspondientes o las interfaces, utilizadas o facilitadas al abonado o cliente, y los datos relativos a la validación del uso del servicio, excluyendo las contraseñas u otros medios de autentificación utilizados en lugar de una contraseña que hayan sido facilitados por el usuario o creados a petición del usuario.
- **Datos relativos al acceso:** datos relativos al inicio y final de una sesión de acceso del usuario a un servicio, que sean estrictamente necesarios con el único fin de identificar al usuario del servicio, tales como la fecha y hora del acceso, o de conexión y desconexión al servicio, junto con la dirección IP asignada al usuario por el proveedor de servicios de acceso a internet, los datos identificativos de la interfaz utilizada y la identificación del usuario. Esto incluye los metadatos de comunicaciones electrónicas según se definen en el artículo 4, apartado 3, letra g), del [Reglamento relativo al respeto de la vida privada y la protección de los datos personales en el sector de las comunicaciones electrónicas].

- **Datos de transacciones**: datos sobre transacciones relacionadas con la prestación de un servicio ofrecido por un proveedor de servicios que sirvan para facilitar información contextual o adicional sobre dicho servicio y sean generados o tratados por un sistema de información del proveedor de servicios, tales como el origen y destino de un mensaje u otro tipo de interacción, la ubicación del dispositivo, la fecha, la hora, la duración, el tamaño, la ruta, el formato, el protocolo utilizado y el tipo de compresión, a menos que estos datos constituyan datos relativos al acceso. Se incluyen aquí los metadatos de las comunicaciones electrónicas, según se definen en el artículo 4, apartado 3, letra g), del [Reglamento relativo al respeto de la vida privada y la protección de los datos personales en el sector de las comunicaciones electrónicas].
- **Datos de contenido**: todo dato almacenado en formato digital, como texto, voz, vídeos, imágenes y sonidos, distintos de los datos de los abonados, los datos relativos al acceso o los datos de transacciones.



4.- FUENTES ABIERTAS: INVESTIGACIÓN EN WEB Y REDES SOCIALES

FUENTES ABIERTAS: MODALIDADES

1. **Redes Sociales** (se examina posteriormente)
2. **Motores de Búsqueda**
   - Los buscadores generales, tales como Google, Bing, Yahoo, Ask o DuckDuckGo
   - Metabuscadores (buscador de buscadores), buscadores personalizados, buscadores especializados y herramientas de recolección de metadatos
   - Los motores de búsqueda proporcionan una facilitación de acceso a la información que es pública, al haber sido aportada voluntariamente para su publicación o difusión en internet, por lo que pueden ser utilizados sin necesidad de autorización judicial por parte de investigadores públicos (agentes policiales) y privados
   - La aplicación de inteligencia a los datos así obtenidos no afecta por sí misma a derechos fundamentales, sin perjuicio de que ello puede determinar operaciones de tratamiento de datos personales
3. **Páginas y sitios web**
   - Libre acceso/acceso restringido a grupo de personas
   - Prueba de una página web: llamados Terceros de Confianza o Prestadores de Servicios de Confianza Digital («Trusted Third Party»-TTP)

FUENTES ABIERTAS: MODALIDADES

4. **Otros**:
   - Datos de personas físicas y de entidades
   - Identificación de propietarios de nombres de dominio (WHOIS)
   - Rastreos informáticos en Redes P2P
5. **Deep Web**: contenido de Internet que no es indexado por los motores de búsqueda tradicionales, o bien solamente se puede acceder a ella mediante el uso de software especializado
   - Dark net: parte de la deep web, donde todo es anónimo y está cifrado siempre , de tal manera que solamente se puede acceder al mismo a través de un programa o protocolo específico (Tor-The Onion Router , I2P...) .

(Imagen obtenida de http://www.grafix.es/internet-profund-o-deep-web-no-tot-el-contingut-es-accessible/)

Inteligencia en fuentes abiertas-OSINT (Open Source Intelligence)

- selección y adquisición de la información, así como un posterior procesado y análisis de la misma con el fin de obtener conocimiento útil y aplicable en distintos ámbitos (en este caso en la investigación y prueba del delito)
- información puede proceder de fuentes abiertas muy heterogéneas: medios de comunicación: revistas, periódicos, radio, etc.; información pública de fuentes gubernamentales; oros, redes sociales, blogs, wikis, etc; conferencias, simposios, «papers», bibliotecas online, etc. Y podrá ser objeto de análisis junto con datos que pueden haberse obtenido de otros medios de investigación de naturaleza tecnológica (balizas, registro de dispositivos…) o tradicionales (vigilancias y seguimientos….).
- no afecta a ningún derecho fundamental, siempre y cuando todos los datos objeto del análisis hayan sido obtenidos de forma legítima y su tratamiento se realice de conformidad con la normativa sobre protección de datos personales

RETO: IMPACTO DE LA IA SOBRE LA PRUEBA DIGITAL

JOAQUÍN DELGADO MARTÍN.

Magistrado Sala Penal Audiencia Nacional (AN). Miembro Red Judicial Expertos Derecho Europeo (REDUE). Doctor en Derecho

¿Cómo afecta la IA a la prueba a la prueba digital?

En un contexto en el que la eficacia probatoria (fiabilidad y licitud) de las pruebas digitales no está siendo fácil de afrontar por los jueces, la IA aporta nuevos elementos de complejidad que se concretan principalmente en:

- Mayores peligros sobre la autenticidad y fiabilidad de las pruebas: alteración de las pruebas y deepfakes (manipulación total o parcial de la prueba)
- Ausencia de una regulación y/o directrices (recomendaciones) sobre cómo afrontar la prueba digital generada por la IA

Esquema general

1. IA para elaborar pruebas falsas (deepfakes)
2. IA para mejorar la calidad de las pruebas
3. IA para generar pruebas a partir de simulaciones
4. IA para investigar hechos ilícitos (genera fuentes de prueba)

1- IA para elaborar pruebas falsas (deepfakes)

¿Qué es una deepfake?

1. “Deep” (creación mediante deep learning o aprendizaje profundo) + “Fake” (falso)
2. Es un contenido de imagen, audio o vídeo generado o manipulado por una IA que se asemeja a personas, objetos, lugares, entidades o sucesos reales y que puede inducir a una persona a pensar erróneamente que son auténticos o verídicos (definición de ultrasuplantación del art. 3.60 RIA)

¿Cómo accede la deepfake al proceso?

- **Fuente de la prueba** = sistema IA que crea o elabora la deepfake
- La deepfake (output o resultado del sistema IA) es un **documento electrónico** (texto, audio, imagen o video).
- **Medio probatorio** = accede al proceso mediante la prueba digital
- Se aplica todo el régimen jurídico de la **prueba digital**.
- Uso de la IA para generar prueba falsa = aporta una **mayor complejidad**
  - a la función del juez para valorar su eficacia probatoria;
  - y la actividad de las partes relativa a la alegación y acreditación de la manipulación total o parcial de la prueba mediante IA (deepfakes)

¿Qué peligros generan en el proceso judicial?

1. La propia falsedad de los datos tenidos en cuenta por el juez para declarar probados los hechos relevantes del proceso.
   - Dificultades de detección de la falsedad
   - Continua sofisticación de las técnicas de generación de deepfakes: complica la detección de las manipulaciones
2. El simple conocimiento de la posibilidad de deepfakes lleve a las partes (y a la sociedad que asiste al juicio de conformidad con el principio de publicidad) a creer que una imagen genuina es falsa.
   - "Dividendo del Mentiroso“: surge cuando la capacidad de crear falsificaciones convincentes permite a los creadores de imágenes socavar la veracidad de información real al afirmar que esta también es una fabricación

¿Cuáles son los efectos jurídicos de la aportación de deepfakes como pruebas?

- **Dimensión procesal**:
  - La prueba no desplegará efectos en el proceso
- **Dimensión penal**:
  - La acción sería constitutiva de delito de **estafa procesal**
  - Relación con el delito de falsedad documental: concursos
  - Sin perjuicio de que pueda ser cometido algún delito en relación la propia elaboración de la deepfake: por ejemplo, si atenta contra la **integridad moral** de la persona afectada

¿Cómo identificar una deepfake en el proceso?

1. Consejos o pistas para la identificación
2. Herramientas técnicas ofrecidas en la web
3. Técnicas especializadas utilizadas por peritos informáticas

Consejos o pistas para identificar el posible contenido falso generado con IA (INCIBE)

Analiza la calidad del contenido: presta atención a ciertos detalles para desenmascararlos en función del tipo de contenido:

- Texto: Coherencia y gramática: verifica si el texto tiene errores gramaticales o incoherencias que no son comunes en artículos escritos por humanos.
- Vídeo:
  - Parpadeos y movimientos faciales: los deepfakes a menudo presentan un número anormal de parpadeos o movimientos faciales poco naturales
  - Incongruencias entre el rostro y el cuerpo: observa si hay desajustes entre la expresión facial y los movimientos corporales
  - Calidad del sonido: el sonido puede no concordar con la imagen, mostrando desfases o variaciones en el tono que no se corresponden con la escena

Consejos o pistas para identificar el posible contenido falso generado con IA (INCIBE)

- Audio:
  - Tono y ritmo: analiza si el tono de voz y el ritmo del habla son consistentes. Las clonaciones de voz pueden presentar variaciones inusuales.
  - Ruido de fondo: la falta de ruido de fondo o la presencia de sonidos extraños puede indicar manipulación.
  - Contexto del discurso: verifica si el contenido del discurso es consistente con lo que la persona normalmente diría o en el contexto en el que suele hablar.
- Imagen:
  - Detalles finos: pueden tener fallos en ciertos detalles, como manos con dedos desproporcionados o fondos que no se alinean correctamente.
  - Iluminación y sombras: revisa si la iluminación y las sombras son coherentes en toda la imagen.

Herramientas técnicas útiles para identificar deepfake (INCIBE)

1. Detectores de IA para texto: Existen varias herramientas diseñadas para analizar textos y detectar si han sido generados artificialmente. Mostramos algunos ejemplos:
   - plagiarismdetector: analiza patrones lingüísticos y estructuras gramaticales para indicar un tanto por ciento de posibilidades de que sea generado artificialmente.
   - GPTZero: tiene la capacidad de detectar contenido generado por modelos como ChatGPT y GPT-4. Muestra los porcentajes de probabilidad de autoría humana o IA, y resalta las secciones sospechosas.
   - 3Copyleaks: proporciona un análisis detallado a nivel de oración para detectar contenido generado por IA con una alta precisión.

Herramientas técnicas útiles para identificar deepfake (INCIBE)

2. Herramientas de detección de multimedia: Para detectar deepfakes y otros tipos de contenido multimedia generado por IA, estas herramientas pueden ser muy útiles:
   - VerifAI, una web respaldada por la empresa Telefónica que detecta en pocos segundos si un contenido ha sido generado o alterado con IA.
   - Resemble Detect detecta audios generados con IA.
   - Deepware: permite analizar vídeos y detectar deepfakes mediante un análisis exhaustivo que identifica alteraciones y anomalías.
   - Illuminarty: proporciona análisis avanzado para detectar deepfakes en imágenes, utilizando inteligencia artificial para identificar discrepancias y manipulaciones.
   - AI or Not: permite a los usuarios cargar imágenes y audios para verificar si han sido manipulados o generados por inteligencia artificial.
   - VerificAudio: ayuda a verificar, enviando a un equipo especializado, voces sintéticas.

Conclusiones

Frente a la mayor complejidad de la actividad probatoria ligada a la posible utilización de deepfakes

- Resulta necesaria una mejor formación de los jueces , fiscales y abogados sobre el posible uso de IA para la manipulación de las pruebas .
- Y adquiere una mayor relevancia de la prueba pericial informática para la acreditación y/o impugnación de la fiabilidad (autenticidad e integridad) del documento electrónico aportado al proceso
  - Necesaria la capacitación y preparación del perito en esta materia, específicamente en relación con instrumentos o técnicas de detección de deepfake: la misma tecnología que se utiliza para generar a las deepfakes (redes neuronales generativas adversariales); análisis de compresión del vídeo; análisis de los metadatos del archivo….
- La posibilidad de existencia de deepfakes puede determinar un incremento del coste económico del proceso, así como una extensión de su duración.

2- IA para mejorar pruebas

IA para mejorar la calidad de las pruebas

- "Pruebas digitalmente mejoradas": audios, videos o imágenes que han sido mejorados mediante software de inteligencia artificial
- Posibilidades:
  - Mejorar la percepción del destinatario (juez), como ocurre con acercar una imagen, acelerar o desacelerar un video, o separar una voz del ruido de fondo;
  - Completar la imagen (rellenando píxeles) o el audio con lo que el software IA considera que debería estar allí, alterando el original.
- No afecta a la eficacia probatoria, siempre que resulta acreditado que no se ha modificado la prueba

3- IA para generar pruebas a partir de simulaciones

IA para generar pruebas a partir de simulaciones

- IA puede crear simulaciones basadas en datos recopilados durante la investigación
  - Investigación pública: por agentes policiales en el proceso penal;
  - Investigación privada, aplicada por las partes para aportar pruebas en procesos ante cualquier jurisdicción.
- Ejemplos:
  - Reconstrucción de escenas del crimen: simulaciones 3D que muestran los eventos según la evidencia recopilada; y
  - Análisis de trayectorias: la simulación de trayectorias de disparos o movimientos de los involucrados en un crimen.

4- IA para la investigación de actos ilícitos

IA para investigar actos ilícitos

- Técnicas IA que usan datos biométricos
  - Reconocimiento facial
  - Reconocimiento de voz
  - Reconocimiento de emociones
  - Reconocimiento de huellas dactilares
  - Reconocimiento de ADN
  - Reconocimiento de firma y escritura
- Técnicas IA que emplean técnicas de Visión Artificial (Computer Vision)
  - Análisis de imágenes
  - Lectura de matrículas
  - Detección de documentos falsos
- Otras técnicas IA
  - Detección de estafas o fraudes digitales.
  - Detección de disparos
  - Detección de contenidos ilícitos en plataformas digitales