IA en la seguridad y delincuencia

Implicaciones político-criminales para el futuro

David Pellejero Cuenca

Directora: Dévika Pérez Medina

Máster en Ciberdelincuencia

M5.258 Trabajo Final de Máster

Curso académico: 2022-2023

Fecha: 15 de Junio de 2023

RESUMEN: La Inteligencia Artificial (IA) es un fenómeno en constante evolución que ha revolucionado la forma en la que el ser humano interactúa con la tecnología. Desde su concepción teórica inicial, la IA ha ido evolucionando hasta convertirse en una realidad 'tangible' hoy en día. Como todos los avances tecnológicos destacables, este fenómeno también ha dado lugar a nuevas adaptaciones de comportamientos delictivos, planteando nuevos desafíos para la seguridad y la aplicación de la ley, cuya respuesta ha sido amoldarse a dicha tecnología disruptiva. El objetivo de este trabajo es analizar la jurisprudencia europea propuesta en materia de inteligencia artificial y derechos fundamentales, poniéndola en relación con los fenómenos delictivos que se puedan derivar de su uso y la presunta adaptación por parte de los Cuerpos y Fuerzas de seguridad. A este respecto, el resultado del estudio muestra la necesidad de ampliar y mejorar la regulación propuesta en cuanto al uso legal de 'sistemas de IA' y la revisión de ciertos presupuestos penales que se podrían adaptar a este nuevo medio comisivo, a fin de alcanzar una efectiva protección de los derechos fundamentales de las personas.

Palabras Clave: Inteligencia Artificial; ciberespacio; análisis criminológico; derechos fundamentales

ABSTRACT: Artificial Intelligence (AI) is an ever-evolving phenomenon that has revolutionized the way humans interact with technology. From its initial theoretical conception, AI has evolved to become a 'tangible' reality nowadays. Like any significant technological advancement, this phenomenon has also given rise to new forms of criminal behaviour, presenting new challenges for security and law enforcement, which in turn require adaptation to this disruptive technology. The objective of this work is to analyse the proposed European jurisprudence regarding Artificial Intelligence and fundamental rights, and to relate it to the criminal phenomena that may arise from its use and the alleged adaptation by law enforcement agents. In this regard, the study's findings highlight the need to expand and improve the proposed regulations regarding the legal use of 'AI systems', as well as the revision of certain legal frameworks that could be adapted to this new modus operandi, in order to effectively protect the fundamental rights of individuals.

Key words: Artificial Intelligence; cyberspace; criminological analysis; fundamental rights.

Índice

1. INTRODUCCIÓN

En la sociedad actual, la Inteligencia Artificial (IA) se ha integrado de forma imperceptible en nuestra vida cotidiana y su desarrollo es cada vez más rápido. En palabras de ALONSO (2021): "está llamada a transformar nuestras vidas, incidiendo en muchos campos distintos como el de la medicina, la agricultura, la educación", así como los campos relevantes para este estudio, la criminología y el derecho. Teniendo en cuenta su crecimiento, en constante evolución, y su interacción con tantas áreas que afectan a la ciudadanía, su uso puede llegar a afectar múltiples derechos fundamentales. Por lo tanto, será importante el análisis de las implicaciones de la IA en materia preventivo-penal, tanto en los usos presentes y futuros, como en los posibles delitos emergentes derivados de su utilización.

Para ello, este trabajo está estructurado en dos bloques. En primer lugar, se describe el contexto teórico en el que se centrará el trabajo, incluyendo una explicación y descripción del concepto de forma técnica y legislativa, para la posterior discusión, así como las innovaciones procedimentales que ha implicado o está implicando esta tecnología disruptiva. El otro gran eje del trabajo está conformado por la parte de índole más práctica, dónde haciendo uso de tablas de datos, se realizará un análisis doctrinal y normativo sobre el uso de la IA en la seguridad y en la cibercriminalidad.

Por lo tanto, el presente trabajo se propone analizar las implicaciones de la IA en la materia preventivo-penal, abordando tanto sus aspectos teóricos como prácticos. Mediante un enfoque interdisciplinario que involucrará la criminología, la legislación, la ética y la tecnología, se busca comprender la naturaleza de la IA, sus posibles impactos en el sistema de justicia penal, las medidas necesarias para asegurar su uso responsable y acorde a los derechos fundamentales de las personas y las posibles implicaciones de esta tecnología disruptiva en la cibercriminalidad. A través del análisis riguroso y fundamentado, se espera contribuir al debate y vislumbrar si la legislación venidera está realmente preparada para comprender y regular la complejidad de la Inteligencia Artificial.

2. HIPÓTESIS Y OBJETIVOS

Este estudio parte de la siguiente pregunta de investigación: '¿Qué implicaciones tiene la Inteligencia Artificial para la ciberdelincuencia, la seguridad y el derecho?'. Por lo tanto, este trabajo tiene como objetivo principal la visibilización y divulgación de la problemática que puede suponer la utilización de la IA, tanto para la seguridad como para la ciberdelincuencia, comparándola con la legislación existente y la que está por venir. Para ello, este objetivo será dividido en diferentes objetivos secundarios, en primer lugar, alcanzar el resultado del análisis del estado de la cuestión acerca del concepto de la IA que a su vez se abordará desde la dimensión de la cibercriminalidad y desde la seguridad; en segundo lugar, se ha realizado un análisis detallado de diferentes herramientas usadas tanto en la investigación penal y procesal como en la ciberdelincuencia, abordando diferentes características importantes de las mismas como el tipo de aprendizaje de la IA, el riesgo hacía los derechos fundamentales de las personas, la trazabilidad de dichas técnicas, el tipo de uso que se le da a la IA, y otros elementos expuestos en su apartado correspondiente; por último, la unión de estos dos primeros bloques, en un apartado de resultados, que responde a la pregunta de investigación. Además, con tal de complementar los objetivos secundarios anteriormente comentados y colaborar en la resolución de la pregunta de investigación, este estudio se plantea las siguientes hipótesis:

1- La IA se utiliza de manera complementaria para cometer delitos.
2- El uso de IA por parte de la investigación criminal no contempla todas las posibles implicaciones negativas de su utilización.
3- El uso de IA por parte de la seguridad implica igual o mayor riesgo para los derechos de las personas que el uso por cibercriminales.
4- No será necesaria una actualización del derecho penal para nuevas modalidades delictivas que puedan surgir gracias a la IA.

Asimismo, es esencial recalcar que esta investigación tiene como fin la obtención de información relevante para el futuro de la IA en el campo de la criminología y el derecho penal, y no se plantea generalizar la problemática planteada a todos los campos en los que es de aplicación la IA, sino preparar el terreno para futuras investigaciones en los campos mencionados.

3. METODOLOGÍA

Este trabajo es un estudio criminológico y jurídico, que de acuerdo con las técnicas que se van a utilizar para resolver la casuística de este, es de carácter cualitativo. Por lo tanto, siguiendo lo expuesto en PETZOLD-PERNÍA (2008), los razonamientos empleados generarán datos de carácter jurídico que serán guiados por las leyes pertenecientes a Europa y, de manera más concreta, a España.

Para la elaboración de las definiciones y conceptos desde una perspectiva técnica, se han utilizado las bases de datos de *Google Scholar, Dialnet, Scopus* y *Web of Science*, así como los recursos disponibles en la biblioteca online de la UOC. A fin de obtener información lo más actualizada posible, se ha dado prioridad a los artículos cuya publicación fuera después del 2019, debido a que los 'sistemas de IA' son una tecnología que está en constante desarrollo. Además de esta limitación temporal, también se han utilizado palabras clave, en inglés y español, para acotar la búsqueda de manera más específica, como 'Inteligencia Artificial / Artificial Intelligence', 'Delitos por IA / AI crimes', 'Seguridad e IA / Security AI', 'IA y Derecho / AI and legislation', entre otras, obteniendo como resultado final cuarenta documentos. Además de los artículos científicos consultados, también se han utilizado manuales criminológicos especializados en cibercrimen, como Miró (2012), Holt, Bossler y Seigfried-Spellar (2018) y Agustina, Montiel y Gámez-Guadix (2020). En cuanto la perspectiva jurídico-legislativa, se ha utilizado información incluida en la propuesta de Reglamento sobre Inteligencia Artificial y el Libro Blanco sobre Inteligencia Artificial en relación con otros reglamentos como el RGPD.

Para la resolución del segundo bloque del trabajo, se han desarrollado tablas informativas con los datos recopilados sobre los 'sistemas de IA' analizados. Dicha información, ha sido obtenida tanto en las páginas web de los desarrolladores como de los documentos técnicos, en caso de estar publicados, y los artículos científicos relacionados con dichas herramientas. En lo que respecta la parte jurisprudencial, se han utilizado mayoritariamente, el motor de búsqueda del Consejo General del Poder Judicial (CENDOJ) y el portal de acceso al Derecho de la Unión Europea (EurLex), así como el Código Penal español para el análisis normativo. Igual que para la búsqueda

teórica, se han utilizado términos relacionados con el estudio como ‘Inteligencia Artificial / Artificial Intelligence’, ‘Deep Fakes’, entre otros.

4. MARCO TEÓRICO

4.1. Evolución del concepto de Inteligencia Artificial desde la perspectiva técnica

4.1.1. Primeras aproximaciones de Inteligencia Artificial

Para hablar del concepto de Inteligencia Artificial, su definición y sus tipos, es importante remontarse a su origen. De manera anterior al nacimiento del concepto de Inteligencia Artificial, es importante remontarse al 1936, dónde Alan Turing publicó un artículo sobre los números computables y sus aplicaciones, en dichas aplicaciones fue dónde se inició la teorización para los ordenadores, y, mucho más importante para este trabajo, dónde se empezó a conceptualizar la automatización de los procesos computacionales que unos años más tarde se usarían para la creación de las primeras IA (Turing, 1936).

Otro de los puntos teóricos clave tuvo lugar en 1943 de la mano de WARREN MCCULLOUGH y WALTER PITTS, que desarrollaron el primer modelo matemático para la creación de una red neuronal, sentando las bases para el desarrollo artificial de ‘una mente pensante’. Unos años después, en 1950, Alan Turing publicó otro artículo dónde se introducía el ‘Test de Turing’, una herramienta capaz de evaluar si una máquina podía mostrar comportamientos inteligentes similares a los de un ser humano (Gupta *et al.,* 2021). Con estas publicaciones y otros hechos, de carácter más práctico, como el primer software que aprendió a jugar ajedrez de manera autónoma (Samuel, 1952; citado en Gupta *et al.,* 2021), se usó por primera vez el término de IA aplicado a un nuevo campo científico en la conferencia de Darmouth por parte de John McCarthy en 1956 (Delgado, 1996-7). En dicha conferencia, se instauró el objetivo de imitar las capacidades del cerebro humano mediante algoritmos, siguiendo la premisa de que los programas informáticos podrían funcionar análogamente a los estados cerebrales. Una vez planteado el objetivo principal de este novedoso campo científico, se pudo definir el campo de la IA como el área de las ciencias computacionales que se ocupa de

desarrollar programas informáticos capaces de ejecutar operaciones comparables a las realizadas por la mente humana, como el aprendizaje o el razonamiento lógico (Alonso y Bolón, 2020)

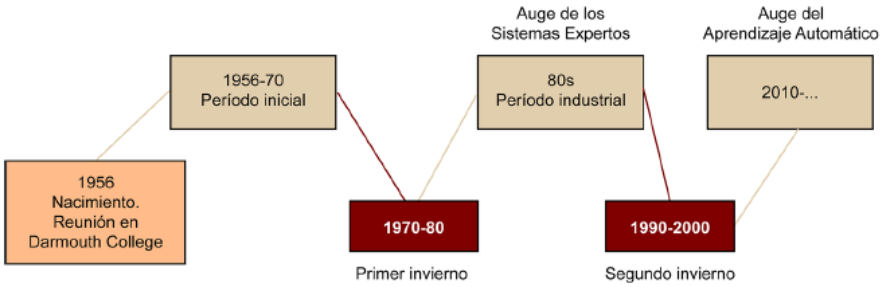
A raíz de esta concepción basada en los algoritmos, aparecieron dos líneas de pensamiento: los que pensaban que estos algoritmos se tenían que fundamentar en la lógica y aquellos que optaban por el uso de la semántica. Gracias a estas dos corrientes pudieron venir las dos etapas más brillantes de la IA en aquellos tiempos. La nueva disciplina de la IA vivía un gran momento, se producían muchos avances en el campo teórico que avalaban ambas líneas de pensamiento, pero como muchos otros campos de investigación, unos años después de su nacimiento aparecieron las primeras críticas a estos sistemas, siendo la principal la reducida escalabilidad de los sistemas de IA planteados a problemas reales, debido a la baja capacidad de procesado de los ordenadores de esa época. Se empezó a reducir la financiación de la investigación de la IA en todo el mundo, entrando en un período triste para este campo, denominado primer invierno, que duró una década (Alonso y Bolón, 2020).

Ya en los años ochenta, surgió un tipo de sistema de IA que parecía solucionar los problemas de escalabilidad anteriores, ya que se priorizó la especialización para poder resolver problemas reales concretos, en vez de intentar diseñar una máquina que fuera capaz de resolver múltiples problemas diferentes. Dichos sistemas fueron denominados ‘sistemas expertos’, cuyo propósito es emular la habilidad de resolución de problemas de un humano en un campo específico. Este logró fue parte de la segunda corriente de pensamiento algorítmica mencionada, es decir, aquella que era más simbólica y no tan matemática. Pese a este gran avance en el campo de la IA, las dificultades que planteaban dichos sistemas en cuanto su mantenimiento, envió a la IA de vuelta al ‘olvido’ durante bastantes años (Alonso y Bolón, 2020).

Pasaron un par de décadas dónde el campo de la IA permaneció en el frío invierno, hasta que, en 2010, la IA, está vez de la mano del pensamiento lógico, volvió a resurgir de sus cenizas gracias a varios factores. El primero fue el ‘Big Data’, la existencia de una infinidad de datos con los que alimentar los sistemas de IA para poder entender cómo evoluciona el entorno; el segundo, y no por ello menos importante, un

mundo cada vez más tecnológico e interconectado; el tercero, relacionado directamente con uno de los motivos que llevó a este campo al desconocimiento, el aumento de la capacidad computacional de los ordenadores pudiendo crear así sistemas de IA más complejos; el penúltimo, avances en *software*, habilitando multitud de programas y herramientas para tratar datos y conseguir plasmar en programas los desarrollos teóricos; y, por último, un aumento de la demanda real de las empresas, a mayor cantidad de datos que se necesitan procesar mayor atracción generará un sistema que realice dicha tarea de la forma más automática posible (Alonso y Bolón, 2020). Fue por todo esto que la inversión económica en materia de IA se disparó de manera abrupta y hoy en día continúa siendo muy importante.

Figura 1. Línea de tiempo del desarrollo de la disciplina de inteligencia artificial.



Fuente: Extraída de "Inteligencia artificial, algoritmos y derecho. Una introducción", por Alonso y Bolón, 2020, *Universitat Oberta de Catalunya*

Una vez vista la evolución del campo de la IA hasta el momento, toca hablar ahora de la percepción más actual del concepto, definiéndolo y explicando los diferentes tipos de sistemas de IA existentes en la actualidad.

4.1.2. Inteligencia Artificial en la actualidad

Primeramente, se abarcará el concepto desde una perspectiva tecnológica, posteriormente se definirá lo que se entiende por IA por parte de las ciencias sociales, en su mayoría desde el derecho, para así poder comparar ambas perspectivas y observar si las dos pueden considerarse sinónimas.

Son muchos los autores que proponen definiciones distintas para el concepto de IA, en función de la utilidad que se le quiere dar, de la definición previa de 'inteligencia' que usan, y muchos otros factores analizados en Wang (2019) que acaban con la siguiente conclusión: "Basándose en este análisis, no existe una definición *correcta* de IA, ya que cada una de ellas tiene valores teóricos y prácticos, por lo que no están *mal*". Partiendo de esta base, dónde no existe una definición global del concepto, debemos tener claro que la misma dependerá del campo de estudio dónde se vaya a usar junto con sus objetivos, aplicaciones y métodos (Wang, 2019). Independientemente de esto, por el bien de este trabajo, se partirá como base la definición de Haenlein y Kaplan (2019): "La IA, es la habilidad de un sistema para interpretar datos externos de manera correcta, aprender de estos y usar dicho aprendizaje para lograr tareas y metas específicas a través de una adaptación flexible". A partir de esta definición, pueden surgir varias preguntas relevantes, como por ejemplo de dónde vienen esos datos y como afectan a los derechos de las personas, pero en este punto nos centraremos en la referente a como aprenden de los datos externos, ya que es dicho aprendizaje previo el que lleva a su posterior utilización.

4.1.2.1. El aprendizaje de la IA

Esta cuestión se ha conceptualizado como 'Machine Learning' (ML), usado este término por primera vez en 1959 por Arthur Samuel (Gupta *et al.*, 2021). Se explicarán a continuación, a partir de Sarker (2021b), los diferentes tipos de algoritmos existentes y ejemplos de uso, para ayudar a entender el proceso de aprendizaje:

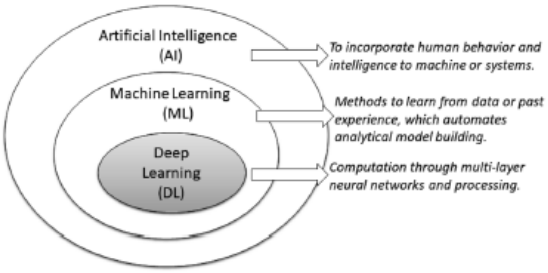
- **Aprendizaje supervisado**: Este algoritmo de aprendizaje utiliza datos de entrada clasificados de manera previa, para que la máquina identifique patrones para poder clasificar los datos en la fase de entrenamiento. Por ejemplo, se le aportan a la máquina fotos de animales con sus etiquetas correspondientes, p. ej. Foto de gato con la etiqueta "gato", a través de su programación la IA identifica patrones relacionados con cada etiqueta específica. Una vez realizada esta fase de aprendizaje se le dan fotos de animales sin etiquetar para que la IA etiquete de acuerdo con los patrones previamente establecidos.

- **Aprendizaje no supervisado:** Sin la ayuda previa de un humano, la IA analiza un conjunto de datos para determinar patrones, estructuras o relaciones en dicha información. Se tiende a usar este tipo de aprendizaje si se quiere observar la relación entre distintos datos, por ejemplo, se le da a la IA diferente tipos de objetos con distintos colores y tamaños, esta aprende que características parecidas tienen esos objetos y los agrupa según esa información, ya sea tamaño, color, u otro tipo de relaciones.
- **Aprendizaje semi-supervisado:** Es un híbrido entre los dos tipos algoritmos anteriormente explicados, se le ofrecen a la IA datos tanto clasificados como sin clasificar, para que esta aprenda a clasificar los que no van previamente etiquetados usando la información disponible en los etiquetados. Un ejemplo claro es el entrenamiento de una IA para clasificar correos spam, se dan varios correos etiquetados como spam y otros que no lo están, dónde suelen ser la mayoría. La IA aprende a clasificarlos según las características de los correos etiquetados previamente.
- **Aprendizaje por refuerzo**: Este tipo de algoritmo de aprendizaje habilitan a la IA para evaluar de manera óptima y automática el comportamiento adecuado según el contexto específico, mediante un elemento caracterizado como refuerzo cuando se realiza la acción adecuada. Un ejemplo claro de esto son las IA entrenadas para jugar al ajedrez, dónde el refuerzo sería ganar la partida. Mediante prueba y error, la máquina analiza los movimientos que le han llevado a la victoria para así almacenar esos datos y poder usarlo en partidas posteriores.

Estos son los tipos básicos de algoritmos de ML existentes, sin embargo, hoy en día, las grandes IA utilizan sistemas mucho más complejos y avanzados de aprendizaje conocido como Deep Learning (DL), debido a sus capacidades de aprendizaje a partir de los datos proporcionados (Gupta *et al.*, 2021; Sarker, 2021a). En 2006, el 'Deep Learning' fue introducido por Hinton et al. basándose en el concepto de las redes neuronales artificiales (citado en Sarker, 2021a). Este nuevo modelo de aprendizaje mejorado se diferencia de los anteriores algoritmos de ML por la eficiencia que aporta debido al aumento en el volumen de datos a gestionar, aunque la tecnología DL requiere de un mayor tiempo de entrenamiento, debido al mayor número de parámetros que se

deben gestionar, requiere un menor tiempo de testeo comparado con los otros algoritmos ML (Saker, 2021a). Aun existiendo esta diferenciación con los modelos de ML presentados anteriormente, el DL con relación a los sistemas de ML se representaría de la siguiente manera:

Figura 2. Ilustración representativa de dónde se sitúa el DL en relación con el ML y a la IA.



Fuente: Extraída de "Deep learning: A comprehensive overview on Techniques, Taxonomy, Applications and Research Directions", por Saker, 2021a, *SN Computer Science,* 2, 420. https://doi.org/10.1007/s42979-021-00815-1

Esta tecnología hace uso de redes neuronales con múltiples capas para entrenar modelos computacionales capaces de aprender representaciones de datos de manera jerárquica y automática, destacan entre dichas capas las capas ocultas, que tienen en su interior algoritmos y programas con la función de segregar mejor la información analizada. Es debido a esta característica del DL que comúnmente se les considera como "black-box machine", ya que, dificultan la investigación de sus procedimientos al sacrificar transparencia y capacidad interpretativa a cambio de precisión en la predicción (Saker, 2021a y Rai, 2019). Dentro del DL existen muchas maneras distintas de clasificar las técnicas existentes, sin embargo, este estudio se basará en la clasificación de Saker (2021a), al fundamentarse en las técnicas ML explicadas anteriormente y al añadir ejemplos más concretos de cada tipo de técnica DL. Por lo tanto, tenemos los siguientes tipos:

- **Red profunda para el aprendizaje supervisado:** Este tipo de DL añade una función discriminatoria a las aplicaciones supervisadas. Dentro de esta clasificación existen diferentes tipos de arquitectura, siendo los siguientes los más conocidos:

- Perceptrón multicapa (MLP): Este es la arquitectura base de las redes neuronales profundas, una red conectada entre sí que tiene una capa de entrada que recibe los datos, una capa de salida orientada a la toma de decisiones de cara a la salida de datos y una o varias capas ocultas actuando como motor computacional de la red, que añaden los elementos discriminadores que ayudarán a la IA a realizar su tarea.
- Red neuronal convolucional (CNN): Arquitectura especializada en el procesamiento de datos de tipo imagen que puede aprender características de los datos de entrada de forma autónoma, solo siendo necesario la etiquetación de la imagen.
- Red neuronal recurrente (RNN): En otro tipo de arquitectura muy similar a la anterior, caracterizada por su sistema de memoria, es decir, aprende de los datos de entrada, pero su sistema de memoria puede influir en posteriores entradas o salidas de datos en función de la información almacenada de entradas anteriores.

- **Red profunda para el aprendizaje generativo:** Esta categoría de DL es utilizada para encontrar patrones y características complejas en los datos de entrada, sin necesidad de tener etiquetados anteriores. Aprenden del conjunto de datos de entrada, buscando correlaciones y distribuciones en los mismos. De la misma manera que las redes para el aprendizaje supervisado, también se distinguen, en Saker (2021a), algunos tipos concretos de modelos:
  - Red adversaria generativa (GAN): Este tipo de arquitectura funciona mediante el descubrimiento y aprendizaje de patrones en los datos de entrada de manera que pueda generar nuevos ejemplos a partir del conjunto de datos original. Están formadas por dos redes neuronales, por un lado, el generador (G) que creara nuevos datos con las propiedades observadas en los datos originales y el discriminador (D) que predice si la muestra observada proviene de datos reales o del generador. Estas dos redes neuronales se retroalimentan entre sí, mejorándose la una a la otra. Para los generadores de imágenes a partir de texto se suelen utilizar este tipo de arquitecturas.

- Codificador automático (AE): Los auto-codificadores son una técnica de aprendizaje muy popular en la que se usan redes neuronales para aprender representaciones. Esta arquitectura se tiende a utilizar con datos de alta dimensión, por lo que se hace necesario para su análisis la codificación de dicha información. Para ello, el AE, consta de tres partes: el codificador, el código y el decodificador, el primero comprime los datos de entrada y genera el código que posteriormente usará el decodificador para poder reconstruir la entrada.

- **Red profunda para el aprendizaje híbrido y otros enfoques:** Dentro de esta categoría nos encontramos la unión entre un modelo de DL que actúa como base y otro tipo diferente de DL que añade complejidad y especialización para diferentes tipos de utilidades. Dentro de este tipo de red profunda se explican algunos modelos híbridos que unen los modelos anteriormente comentados como el CNN+AE, AE+GAN, pero se incidirá más en la explicación dada para los otros enfoques:
  - Aprendizaje profundo de transferencia (DTL): Este modelo de arquitectura se basa en la información aprendida de otros modelos anteriores para resolver nuevas tareas con el mínimo entrenamiento necesario. Se basan en dos fases, el pre-entrenamiento, dónde adquieren los conocimientos aprendidos por otros modelos, y un paso de ajuste final, dónde se entrenará para resolver la tarea objetivo. Este tipo de modelos son muy útiles hoy en día debido a las altas necesidades de recursos y de datos para crear una adecuada IA que trabaje con algún tipo de arquitectura DL.
  - Aprendizaje por refuerzo profundo (DRL): Al igual que el 'Aprendizaje por refuerzo' visto anteriormente, el DRL une ese algoritmo y lo aplica a una red neuronal profunda en el proceso de aprendizaje. Por lo tanto, se basa en el mismo principio de refuerzo cuando se realiza una acción 'buena' para así aprender que acciones son las adecuadas para cada situación.
- **Procesamiento del lenguaje natural (NLP) y la arquitectura 'Transformer'**: Pese a esta categoría no estar integrada en Saker (2021a), al no ser como tal un tipo de aprendizaje de red profunda, se hace relevante mencionar este tipo de arquitectura aquí debido a sus características 'híbridas'. Creado por Vaswani *et al.*

(2017), esta arquitectura se fundamenta en una modificación de las arquitecturas AE y RNN, añadiendo una nueva función denominada 'Attention' que mejoraba el aprendizaje mediante ese nuevo paso de retención de datos. Este nuevo modelo implantó las bases para el procesamiento del lenguaje natural humano implantado en IA. Un ejemplo claro de tecnología disruptiva que usa esta arquitectura es el ChatGPT.

Una vez examinada la perspectiva técnica de la IA, definido el concepto y expuestos los tipos de aprendizaje principales dentro del campo científico, ahora se presentarán las definiciones existentes en materia de derecho.

4.2.Aproximación jurídico-legislativa del concepto Inteligencia Artificial

En la propuesta de Reglamento sobre Inteligencia Artificial de 2021, también denominada como 'Ley de Inteligencia Artificial', siendo así como se le hará mención en este trabajo, se define explícitamente un 'sistema de IA' de la manera más tecnológicamente neutra posible para que así resista al paso del tiempo de la mejor manera. Con este fin, se plantea que un 'sistema de IA' es:

> "El software que se desarrolla empleando una o varias de las técnicas y estrategias que figuran en el anexo I y que puede, para un conjunto determinado de objetivos definidos por seres humanos, generar información de salida como contenidos, predicciones, recomendaciones o decisiones que influyan en los entornos con los que interactúa." ('Ley de Inteligencia Artificial', 2021).

Siendo las técnicas y estrategias empleadas: "Estrategias de aprendizaje automático, incluidos el aprendizaje supervisado, el no supervisado y el realizado por refuerzo, que emplean una amplia variedad de métodos, entre ellos el aprendizaje profundo" y otros tipos de estrategias computacionales, relacionadas con la programación y las matemáticas. Se observa que la definición dada a 'sistemas de IA' cumple con su objetivo, es una definición general, amplia y que puede llegar a resistir en el tiempo. Viendo esta definición no se intuye ninguna discordancia grave entre las definiciones técnicas y jurídicas, ¿entonces cuál es el problema?

El problema, o problemas como veremos a continuación, reside en la definición de ‘sistemas de IA de alto riesgo’, estipulada en el art. 6 de esta misma propuesta de ley, que dice que un sistema de IA se considerará de alto riesgo cuando el mismo esté destinado a ser utilizado como componente de seguridad de uno de los productos contemplados en la legislación de armonización de la Unión, en resumidas cuentas, productos relacionados con vehículos aéreos, terrestres o acuáticos, juguetes, aparatos destinados para su uso en atmósferas explosivas, equipos radioeléctricos, productos a presión, equipos de protección individual y productos sanitarios, además, si están dentro de uno de los anteriores deberá de pasar una evaluación de la conformidad. También se añaden a los ‘sistemas de IA de alto riesgo’ los relacionados con los ámbitos siguientes: identificación biométrica, gestión y funcionamiento de infraestructuras esenciales, educación y formación profesional, empleo, gestión de los trabajadores ya acceso al autoempleo, acceso y disfrute de servicios públicos y privados esenciales y sus beneficios, asuntos relacionados con la aplicación de la ley, gestión de la migración, el asilo y el control fronterizo y administración de justicia y procesos democráticos (art. 6.2 Propuesta de ‘Ley de Inteligencia Artificial’ y ANEXO II y III de dicha ley).

Una vez definido los ‘sistemas de IA de alto riesgo’ se encuentran varias discrepancias en cuanto a lo que se intenta proteger con esta clasificación. Siguiendo los derechos fundamentales establecidos por Europa, se puede observar como la propuesta de ley intenta proteger los mismos, en cuanto clasifica como ‘sistemas de IA de alto riesgo’ algunos sistemas que afectan a ciertos derechos fundamentales, como, por ejemplo, el derecho de asilo. Sin embargo, pese a los intentos de establecer que el enfoque europeo sobre la Inteligencia Artificial se basará en los valores y derechos fundamentales (Libro Blanco sobre la Inteligencia Artificial, p. 2), se dejan de lado en la propuesta de ley muchos otros derechos fundamentales como podrían ser la protección de datos de carácter personal, el derecho a la no discriminación, el derecho a la igualdad entre hombres y mujeres, los derechos del menor y la protección del medioambiente y de los consumidores (art. 8, 21, 23, 24, 37 y 38 Carta de los Derechos Fundamentales de la Unión Europea), pese a que en el Libro Blanco sobre la Inteligencia Artificial si se haga mención, en el apartado “Riesgos para los derechos fundamentales, especialmente

la protección de los datos personales y de privacidad y la no discriminación", al elevado riesgo que pueden suponer dichos sistemas para estos derechos.

Otro de los problemas que surgen del análisis a esta propuesta de 'Ley de Inteligencia Artificial', es la poca consideración por parte de la Comisión sobre los derechos de la propiedad, concretamente, los de la propiedad intelectual (art. 17 Carta de los Derechos Fundamentales de la Unión Europea), ya que no se han abordado mejoras legislativas respecto los problemas de trazabilidad de los derechos de propiedad intelectual en cuanto los resultados generados por IA que han usado datos protegidos (Resolución del Parlamento Europeo, 20 de octubre de 2020, sobre los derechos de propiedad intelectual para el desarrollo de las tecnologías relativas a la inteligencia artificial).

Por otra parte, hablando de los datos usados durante el entrenamiento de las máquinas de inteligencia artificial, si se hace mención en el art. 29 de la propuesta de 'Ley de Inteligencia Artificial', al seguimiento del RGPD, en cuanto la protección de los datos de carácter personal usados por los 'sistemas de IA de alto riesgo', y, por extensión, la protección de los consumidores y los derechos del menor relativos a sus datos personales, pero de nuevo, volvemos a la misma problemática, que no es otra que la escasez de sistemas de IA que entran en dicha clasificación.

Por último, también es relevante destacar la escasa mención a la protección medioambiental, que pese a ser un derecho fundamental (art. 37 Carta de los Derechos Fundamentales de la Unión Europea) no se tiene en cuenta el posible riesgo hacia el mismo, al menos no de manera premonitoria, para el establecimiento de nuevas tecnologías (Vestri, 2023). También es destacable el hecho de que no solo no se valora la protección medioambiental frente a la cantidad de recursos energéticos que pueden llegar a necesitar ciertos avances tecnológicos, ya que únicamente se advierte en una frase que se debe usar de manera respetuosa con este, sino que se usa como argumento para vender su uso, un ejemplo de ello está en las conclusiones del LIBRO BLANCO SOBRE LA INTELIGENCIA ARTIFICIAL (2020): "La IA también puede contribuir a encontrar soluciones a algunos de los problemas sociales más acuciantes, como la lucha

contra el cambio climático y la degradación medioambiental, los retos relacionados con la sostenibilidad y los cambios demográficos" (p. 30).

Por lo tanto, podemos observar cómo legislativamente hablando, pareciere que el año de diferencia entre la publicación del Libro Blanco sobre la Inteligencia Artificial y la propuesta de Ley sobre la Inteligencia Artificial, hayan afectado a la Comisión en cuanto se han olvidado de la premisa más importante que se hizo en el enfoque europeo sobre la IA, que es, ni más ni menos, que el asentamiento de los valores y derechos fundamentales de la Unión Europea en la implementación de los sistemas de IA, y no solo de algunos específicos, sino de todos aquellos que puedan verse afectados por las características técnicas de esta tecnología disruptiva, sin olvidarse de la protección al medio ambiente, derecho que el propio Libro Blanco menciona únicamente para fomentar el uso de la IA, sin tener en cuenta las posibles afectaciones.

4.3.El traslado de los medios comisivos del crimen a las nuevas tecnologías

Una vez vista la conceptualización existente de la IA, tanto en el campo técnico como en el legislativo, se resolverán las cuestiones teóricas que aproximan el término de la IA con la práctica criminal en el ciberespacio, abarcándolas desde la criminología.

4.3.1. Aproximación teórica a la cibercriminalidad

La conceptualización de cibercriminalidad es un concepto que se remonta varias décadas atrás, cuando la aparición de internet como pilar esencial de las TIC acarreo la creación de un nuevo 'espacio vital' dónde los individuos de todo el planeta podían comunicarse 'sin barreras' y de una manera ágil, en comparación a anteriores vías de comunicación como podían ser las cartas. La aparición de este campo provocó en los seres humanos la necesidad de adaptarse al nuevo medio que se estaba formando, instaurando estos nuevos sistemas de comunicación a sus vidas cotidianas, creando estructuras empresariales basadas en las TIC, y como no, haciendo lo propio los comportamientos delictivos (Miró, 2012).

Fue en este momento, que los criminólogos, encargados de estudiar el fenómeno social del crimen, empezaron a conceptualizar de manera teórica y práctica como el ciberespacio se había convertido en un nuevo terreno para las actividades criminales,

naciendo así una nueva disciplina denominada ‘Cibercriminología’ (Moise, 2020; Miró, 2012; Holt, Bossler y Seigfried-Spellar, 2018). En este punto, son varias las vertientes ideológicas que los criminólogos plantearon a partir del artículo de GRABOSKY (2001), titulado “Virtual Criminality: Old wine in new bottles?”, por un lado, podríamos pensar que la ciberdelincuencia es un campo totalmente nuevo y único, dónde se deberán crear nuevas teorías especialmente diseñadas, ya que las tradicionales no sirven en el ciberespacio y, por otro lado, tenemos a aquellos que la entendían como delincuencia tradicional adaptada al nuevo ámbito que es el ciberespacio, siendo esta última la que pareciera ser la más óptima, en cuanto la criminología nunca ha analizado de manera concreta los comportamientos delictivos en el espacio terrenal, sino más bien, analizó la criminalidad que se podía observar en ese momento, y, por lo tanto, puede adaptarse a los nuevos espacios que podrían aparecer (Miró, 2012; Holt, Bossler y Seigfried-Spellar, 2018).

Atendiendo a esto, MIRÓ (2012), clasifica el comportamiento criminal atendiendo a la incidencia de las TIC, primeramente tenemos los casos en que el ciberespacio en sí mismo ha materializado nuevas conductas delictivas dónde las TIC son la única forma de realización de dichas infracciones (cibercrímenes puros); en otros casos, esta nueva realidad virtual no implica nuevas formas delictivas puras, sino réplicas de comportamientos criminales anteriores llevados a cabo en el ciberespacio (cibercrímenes réplica); y, por último, tenemos aquellos delitos que plantean dificultades relacionadas con la prevención de la difusión de contenidos en el ciberespacio (cibercrímenes de contenido).

Siguiendo este enfoque teórico, son muchos los estudios criminológicos aplicados a la cibercriminalidad, en los últimos años, los que utilizan y adaptan las teorías tradicionales al ciberespacio. Tenemos estudios que aplican la teoría del autocontrol (Holt y Steinmetz, 2020; Partin *et al.*, 2021), otros que aplican la teoría del aprendizaje social (Navarro y Marcum, 2020; Shadmanfaat *et al.*, 2019), también los que utilizan la teoría del control social (Luknar, 2022), sin embargo, la teoría más utilizada por parte de la mayoría de los académicos, ya sea de manera troncal o de manera complementaria, es la teoría de las actividades cotidianas (Carrillo, 2020; Holt,

Leukfeldt y van de Weijer, 2020; Suh, Choe y Park, 2020; Bello y Griffiths, 2020; Hawdon, Parti y Dearden, 2020; Kigerl, 2021; Cook *et al.*, 2023). Esto principalmente se debe a que la criminología ambiental, ha sido de las vertientes ideológicas que más ágil se ha adaptado a la modernidad, ya que la tecnología permite una mayor incidencia de los puntos clave de sus teorías, existe mayor número de contactos entre potenciales autores y posibles víctimas acompañado de una reducción de los guardianes capaces al aún no estar establecidos de manera correcta (Miró, 2012).

Por lo tanto, este trabajo se basa también en estas teorías de la criminología ambiental, debido a la fácil traslación de sus conceptos y puntos claves entre el espacio terrenal y el ciberespacio, al focalizar su análisis en el propio espacio de acción dónde podrá aparecer el delito. Además, como se puede observar de la estructura del trabajo, un enfoque teórico centrado en la prevención es el adecuado en cuanto los resultados están orientados a qué mejoras se pueden realizar en el sistema normativo para mejorar el control formal, y, por lo tanto, mejorar la tercera pata de estas teorías, los guardianes capaces (Miró, 2012; Holt, Bossler y Seigfried-Spellar, 2018).

4.3.2. Adaptación del comportamiento delictivo

Centrando ahora el foco teórico de este estudio en el proceso que lleva a los criminales a delinquir en nuevos territorios, seguimos en la criminología ambiental, pero pasamos de la teoría de las actividades rutinarias a la teoría de la prevención situacional para explicar los distintos motivos que pueden llevar a un desplazamiento del delito. Primeramente, tenemos la siguiente tabla de BARNES (1995; citada por Miró, 2012), que nos muestra los distintos tipos de desplazamiento existentes:

Tabla 1. Los tipos de desplazamiento de la conducta criminal.

Tipo de desplazamiento	*Descripción*
Espacial	Delincuentes que abandonan las zonas donde el delito se ha vuelto más difícil de cometer y comienzan a llevar a cabo actos delictivos en otro lugar.
Objetivo	Delincuentes que abandonan objetivos que están bien protegidos y centran sus esfuerzos en otros más vulnerables.
Temporal	Delincuentes que trasladan a otras horas o días de la semana sus delitos, cuando cometer la infracción es menos arriesgado.
Táctico	Los delincuentes que cambian sus tácticas para evitar aquellos obstáculos destinados a frustrar la comisión del delito.
Autor	Como los delincuentes que suelen cometer ciertos delitos están arrestados o detenidos, otros delincuentes deciden ocupar su lugar.
Tipo de delito	Los delincuentes responden al bloqueo de un determinado tipo de acto delictivo, cometiendo delitos totalmente diferentes.

Fuente: Extraída de "*El cibercrimen: Fenomenología y criminología de la delincuencia en el ciberespacio*", por Miró, 2012.

Se puede observar de la Figura 3, que muchos de estos desplazamientos no tendrían sentido en el ciberespacio per se, pero nos pueden ser útiles para visibilizar la adaptación de los criminales al ciberespacio. Véase como en un ejemplo hipotético dónde en un barrio dónde se podía realizar la actividad correspondiente al tráfico de drogas, se han implantado controles formales eficaces, puede llevar a los individuos que la realizaban a desplazar su actividad a otra zona o podrían llegar a 'desplazar' su actividad al ciberespacio. Cabe destacar, que al haberse desarrollado dichos tipos de desplazamiento en el espacio físico, sería correcto, tal y como menciona MIRÓ (2012), hablar de 'adaptación' en vez de 'desplazamiento', además, siguiendo este hilo conductor, sería adecuado para hablar de la adaptación del crimen en el ciberespacio la utilización de la siguiente tabla de MIRÓ (2012):

Tabla 2. Cuadro de la adaptación del crimen en el ciberespacio.

Adaptación del crimen en el ciberespacio	*Descripción*
De identidad virtual	Los cibercriminales cambian el lugar en el ciberespacio desde el que realizan el ataque o el nombre de la web desde el que actúan criminalmente.
De objetivo	Los cibercriminales desechan el ataque a objetivos bien protegidos y centran sus esfuerzos en otros más vulnerables.
Técnica	El cibercriminal mejora su ataque y utiliza nuevos instrumentos para superar las nuevas barreras.
Tipo de delito	Los delincuentes responden al bloqueo de un determinado tipo de acto delictivo, cometiendo delitos totalmente diferentes.

Fuente: Extraída de "*El cibercrimen: Fenomenología y criminología de la delincuencia en el ciberespacio*", por Miró, 2012.

Con esta información, se puede concretar que tipos de adaptación pueden llegar a realizar los cibercriminales cuando aparecen tecnologías disruptivas que pueden explotar para seguir realizando sus comportamientos delictivos. En relación directa con el uso de la IA en materia delictiva, podríamos observar cómo los dos primeros tipos de adaptación, son intrínsecos a cualquier ciberdelito existente, al ser un cambio de objetivo o de lugar que poco tiene que ver con la técnica o tecnología utilizada para llevarlo a cabo, sin embargo, entrando en los siguientes dos tipos, podemos observar como la adaptación técnica del cibercrimen si tiene una relación directa con el uso de tecnologías disruptivas para realizar delitos, a su vez, será interesante la observación del último tipo de adaptación, en cuanto esta puede ir de la mano con la actualización legislativa de nuevos delitos emergentes, como ya se explicará en los resultados.

4.4.El uso de tecnologías disruptivas en el marco de la investigación penal

A continuación se contempla la adaptación del otro bando, es decir, como la investigación penal ha actualizado sus procedimientos y sus actuaciones a raíz de las nuevas tecnologías en constante cambio.

4.4.1. Adaptación tecnológica de la investigación penal a la delincuencia tradicional

De igual manera que para las oportunidades criminales, las innovaciones tecnológicas también mejoran los sistemas de prevención e investigación criminal existentes y permiten la habilitación de nuevos. Históricamente, se puede observar como las nuevas tecnologías han llevado a las Fuerzas y Cuerpos de seguridad del estado a mejorar sus sistemas de gestión, en cuanto a optimización de los procedimientos, siendo un ejemplo observable la aparición de los sistemas radiofónicos que permitieron una mejor coordinación grupal (Ortega *et al.*, 2021). Unos años más hacia adelante, otras tecnologías novedosas, como el desarrollo de las TIC, implicaron para la investigación penal grandes mejoras en cuanto a gestión de los datos e información, a su vez, los avances en el campo del internet de las cosas (IoT, de sus siglas en internet), permitieron la creación de herramientas muy útiles para las Fuerzas y Cuerpos de seguridad del estado, entre los ejemplos más destacados tenemos los circuitos cerrados de televisión ahora conectados entre sí y al ordenador principal gracias a Internet, los sistemas GPS integrados en dispositivos de tamaño reducido, como son las unidades 2Track (Gobierno de España, s. f.) y otras tecnologías añadidas a dispositivos portátiles que permiten la escucha o monitorización de comportamientos sospechosos (Ortega *et al.*, 2021).

Otro uso novedoso de tecnologías emergentes podría ser la utilización de drones para la vigilancia aérea como mejora frente los satélites y helicópteros usados anteriormente con el mismo propósito, siendo los drones una mejora en cuanto a simplicidad de uso y reducción de costes que permite a los cuerpos de seguridad añadir medidas de control aéreo de manera más asequible (Klauser, 2021).

Acercándonos más al punto central de este trabajo, también es observable como la investigación penal también se ha adaptado a las últimas novedades tecnológicas, añadiendo a sus tradicionales funciones de evaluación de situaciones el uso de herramientas actuariales como los algoritmos o inclusive aquellos con capacidades de aprendizaje como las IA. Hay varios ejemplos de este tipo de herramientas como la generación de mapas delictivos tecnológicos, la capacidad de añadir a los sistemas de

vigilancia, herramientas de reconocimiento biométrico u otros que se verán de manera más extendida en apartados posteriores (Alonso, 2021).

Como se observa de este apartado, la investigación penal también ha optado por actualizar su metodología y procedimientos con la entrada de nuevas tecnologías, sin embargo, de manera contraría a la adaptación criminal, el uso de estos nuevos medios investigativos tiene que estar amparado y aceptado por las leyes y por los derechos fundamentales, cosa que reduce significativamente la proliferación del uso de muchas de las técnicas expuestas.

4.4.2. Prevención tecnológica para el ciberdelito

Tras examinar la incorporación de las tecnologías al campo de la investigación penal en materia de delincuencia no digital, es interesante el análisis sobre las innovaciones de los cuerpos de seguridad en materia de ciberseguridad.

Si bien es cierto que el control social formal aplicado en el ciberespacio es un tema muy complicado de gestionar, en cuanto, choca de manera directa con una de las características extrínsecas de Internet, como lo es su carácter desregulado, y a su vez, podría llegar a chocar con el derecho a la libertad (art. 6 Carta de los Derechos Fundamentales de la Unión Europea), sí es un elemento importante a la hora de analizar el cibercrimen, sobre todo cuando se realiza dicho análisis desde una perspectiva ambiental de la delincuencia dónde una posible institucionalización del control formal serviría como un preventivo y eficiente guardián capaz (Miró, 2012).

Pese a la anterior afirmación, el ciberespacio sigue siendo un lugar donde la autodefensa seguirá siendo la forma óptima de protección (Miró, 2012), cosa que lleva a la ciberseguridad al sector privado, ya sea a través de las empresas de antivirus, el seguimiento de estándares de ciberseguridad como las normas ISO, la necesidad de las empresas de tener su propio equipo de ciberseguridad especializado o la contratación de agentes externos para que auditen sus sistemas de protección contra ciberamenazas, dejando a los cuerpos de seguridad la función persecutoria de los cibercrímenes.

Con tal de ejemplificar que técnicas de investigación penal tecnológica pueden realizar los cuerpos de seguridad, se exponen las reguladas en el Estado español. Las

mismas se pueden dividir en técnicas que no necesitan autoridad judicial y las que si la necesitan (Quevedo, 2017). Empezando por las primeras, tenemos la obtención de direcciones IP; la identificación de IMEI, IMSI y MAC; la obtención de datos desvinculados de los procesos de comunicación, como serían la identificación de titulares de conectividad o el acceso a datos no integrados en un proceso comunicativo; la orden de conservación de datos; la captación de conversaciones públicas; y por último, en aquellos casos de emergencia también se podrá, sin autorización judicial, la interceptación de las comunicaciones telefónicas y telemáticas y el registro de dispositivos de almacenamiento masivo de la información. En cuanto las técnicas que si requieren una autorización judicial tenemos las órdenes en relación con la cesión de datos sobre tráfico almacenado; la interceptación de las comunicaciones telemáticas como vía de investigación criminal de los ilícitos que se cometen a través de la red, ya sean correos electrónicos, SMS, mensajería instantánea y/o comunicaciones VoIP; el registro de dispositivos de almacenamiento masivo de la información; registros remotos sobre equipos informáticos; la utilización de agentes encubiertos informáticos y los hallazgos casuales, obtenidos mediante alguna de las otras técnicas mencionadas anteriormente, que necesitarán de una autorización judicial para aumentar el registro anterior que ha llevado a descubrir dichas pruebas (Quevedo, 2017).

5. CARACTERÍSTICAS DE LOS SISTEMAS DE IA: ANÁLISIS DE CASOS

Después de haber explorado los aspectos teóricos que brindan el contexto necesario para este trabajo, se procede a abordar la parte práctica del mismo. En esta sección, se presentan dos tablas que destacaran diversas características de los sistemas de IA, tanto aquellos en uso actual como posibles casos de uso hipotéticos. Estas tablas se enfocarán en los dos contextos de estudio: seguridad y criminalidad.

En la primera tabla, se analizan y evalúan diferentes características clave que deberían de comprender los sistemas de IA utilizados en el ámbito de la seguridad para actuar con conformidad a la propuesta de ‘Ley de Inteligencia Artificial’. Los aspectos esenciales de esta tabla serán, la trazabilidad, en términos de capacidad de seguimiento de los procesos que comprenden el sistema; conocimiento sobre el origen de los datos, tanto en entrenamiento como en uso real, para saber si las fuentes utilizadas acarrean

sesgos que puedan llevar a la herramienta a realizar discriminaciones sociales; el tipo de aprendizaje de la herramienta; la privacidad, en cuanto protección y gestión adecuada de los datos conforme el RGPD; y, por último, los derechos fundamentales que se pueden vulneran**.** La categorización de estas variables se realiza en función de indicadores cualitativos escalares (alto, medio, bajo), tanto para la variable trazabilidad como para la de privacidad; indicadores dicotómicos (sí/no) para la variable de conocimiento del origen de los datos; e indicadores explicativos utilizando los tipos de aprendizaje expuestos anteriormente en el trabajo y los artículos de la Carta de los Derechos Fundamentales de la Unión Europea.

En cuanto la segunda tabla, orientada a los usos, en el ámbito de la delincuencia, se valorarán las mismas variables, dado que se plantea analizar el sistema de IA utilizado en cada comportamiento y se añadirá la variable, tipo de uso de la IA**.** Para categorizar esta nueva variable, será fundamental la conceptualización de cibercrímenes presente en el "apartado 4.3.1", ya que, se adaptará a la utilización de sistemas de IA en cuanto la utilización conlleve nuevos delitos exclusivos de dichos sistemas (Puros); delitos anteriores, ya sean réplicas de comportamientos delictivos tradicionales como de las adaptaciones informáticas, en los que se utilicen sistemas de IA para llevarlos a cabo (Réplica); y, por último, se seguirán contemplando los ciberdelitos de contenido, pero que utilicen sistemas de IA para su realización (Contenido).

5.1.Ámbito de la seguridad

En cuanto los sistemas de IA que tienen relación con el campo de la seguridad son muchos los posibles candidatos teóricos que podrían plantearse, sin embargo, se opta por hacer una revisión de aquellos que si han tenido, o tienen actualmente, cabida en las Fuerzas y Cuerpos de seguridad (Alonso, 2021). Los dos primeros, son sistemas de IA en uso o en vías de aplicación en el estado español y el tercero es un sistema de IA utilizado en los Estados Unidos.

Tabla 3. Sistemas de IA utilizados en el ámbito de la seguridad.

Sistemas de IA	Trazabilidad	Datos	Aprendizaje	Privacidad	Derechos
VeriPol	Alta	Sí	NLP+ML	Media	Art. 6, 8 y 21
ABIS	Baja	Sí	*ML*	Alta	Art. 8 y 20
PredPol	Baja	No	*ML*	Baja	Art. 8, 20 y 21

Fuente: Elaboración propia a partir de datos extraídos de diferentes fuentes citadas en sus apartados correspondientes.

5.1.1. VeriPol

Los datos analizados sobre el sistema VeriPol, un sistema de IA utilizado en España para detectar denuncias falsas (Alonso, 2021), se han extraído de Quijano, Liberatore, Camacho y Camacho (2018).

Primeramente, en cuanto a la trazabilidad del modelo planteado, se explica de manera correcta el procedimiento por el que pasan los datos y se exponen de manera adecuada su diseño y fiabilidad alcanzada.

En cuanto los datos utilizados para el entrenamiento, se indica que son un conjunto de 1122 denuncias, de las cuales 534 eran verdaderas y 588 eran falsas, también se tiene constancia de los datos que se utilizaron para probar el modelo en un entorno real, siendo estos los casos de robo que llegasen en el área de Málaga y Murcia en un periodo de 4 días en cada ciudad. Respecto los datos utilizados por la herramienta en su uso real serán los datos que se reciban directamente de los testimonios de las víctimas.

El tipo de aprendizaje de la máquina también está definido en el artículo, combinando dos técnicas explicadas en este trabajo, como son el NLP sumado con un aprendizaje supervisado.

Sobre la privacidad, se destaca el uso de técnicas de anonimización de los datos utilizados durante la fase de entrenamiento, pero no se hace ninguna mención a la protección de datos y mucho menos a la LOPDGDD o al RGPD, sin embargo, al tratarse de una herramienta diseñada en conjunto con las Fuerzas y Cuerpos de

seguridad del estado, dónde el RGPD tiene cabida, se presupone, a fin de categorizar esta variable, que sí existe conocimiento sobre problemas que se puedan acarrear debido a la protección de datos.

Por último, la variable referente a los derechos fundamentales es innegable que los derechos de clara mención en este ejemplo serían el derecho a la vida y a la seguridad (art. 6 Carta de los Derechos Fundamentales de la Unión Europea), en cuanto se podría estar transgrediendo este derecho si el sistema de IA emite un falso negativo ante un testimonio real, lo que podría llevar a ulteriores problemas para estas dos esferas. A su vez, repitiendo lo expuesto en el párrafo anterior, si el sistema de IA no sigue el RGPD se estaría transgrediendo el derecho a la protección de datos (art. 8 Carta de los Derechos Fundamentales de la Unión Europea). También es destacable la posible vulneración del derecho a la no discriminación (art. 21 Carta de los Derechos Fundamentales de la Unión Europea), en cuanto el sistema de IA utiliza como indicador de fraudulencia en el testimonio los usos léxico-sintácticos pobres.

5.1.2. Reconocimiento facial biométrico automático (ABIS)

Pese a que son muchos los sistemas de IA programados para llevar a cabo tareas de reconocimiento biométrico, la herramienta analizada debido a que es el sistema automático de identificación facial biométrico que se está probando y desarrollando en España estos últimos años, será el sistema CABIS diseñado por la compañía Thales (Comisión Europea, 2021; González, 2022; Thales, 2022). Además de especificarse este sistema concreto, se analizará en el contexto español siguiendo el documento de la Comisión Europea, ya que esto servirá para el posterior análisis.

Empezando por la variable trazabilidad, tanto el documento analizado como la propia página de la empresa desarrolladora de dicha tecnología (Thales, s.f.), no muestran de manera clara el procedimiento detrás de dicho sistema de IA, sólo tienen vídeos propagandísticos y menciones a la interfaz web del programa, pero no a como este opera en su interior (Thales. S.f.).

En cuanto a los datos que utilizará el sistema ABIS en territorio español, serán aquellas fotografías obtenidas por parte de los cuerpos de seguridad del Estado al

realizar detenciones las que se compararán con las imágenes obtenidas de las cámaras de seguridad.

Respecto el tipo de aprendizaje, no hay una concreta mención ni en Comisión Europea (2021) ni en Thales (s.f.), únicamente se menciona el uso de ML, sin embargo, se puede deducir del funcionamiento de la aplicación que se trata de un sistema supervisado de ML.

En referencia a la privacidad, si se hace mención del RGPD y a la LOPDGDD, en su apartado correspondiente de Comisión Europea (2021). Además, se menciona la imposibilidad legal que tiene España en referencia al uso de los datos biométricos obtenidos, sin consentimiento, por la expedición de carnés de identidad y pasaportes, ya que se estaría vulnerando completamente el RGPD y la LOPDGDD. Además, en Ministerio del Interior (2023), se hace mención como base legitimadora del uso de esta herramienta el art. 11 de la LO 7/2021, de protección de datos personales tratados par afines de prevención, detección, investigación y enjuiciamiento de infracciones penales y de ejecución de sanciones penales.

De nuevo podemos observar que en la variable de los derechos fundamentales**,** se vuelve a hacer incidencia en el art. 8, en referencia a la protección de datos, por obvias razones. Además, se ha incluido como posible derecho vulnerado el derecho a la igualdad ante la ley (art. 20 Carta de los Derechos Fundamentales de la Unión Europea), en cuanto el RGPD, la propuesta de Ley sobre Inteligencia Artificial y otros documentos legales puede ignorarse en casos dónde se habla de derechos de delincuentes.

5.1.3. PredPol

De manera similar al caso anterior, PredPol es un sistema de IA privado utilizado en la ciudad de Oakland, California; Uruguay y Kent, Reino Unido (Eticas Foundation, 2021). Esta herramienta permite identificar zonas de calor delictivas en un mapa. Por lo tanto, este sistema de IA es la tecnificación de las teorías medioambientales que estipulaban *hot spots* en función del nivel de convergencia de delincuentes y víctimas, atendiendo a un aumento de las oportunidades en dicha zona (Balcells, 2020; citado en

Alonso, 2021). En general hay pocos estudios concretos sobre el uso de esta herramienta o que se añadan explicaciones técnicas sobre la misma, posiblemente debido a la legislación americana sobre la propiedad intelectual, de igual manera, se ha utilizado Rubel, Castro y Pham (2021) y PredPol (s. f.).

Como ya se ha mencionado al inicio de este apartado, la trazabilidad es nula, no hay información sobre el procedimiento que siguen los datos, únicamente se muestra en PredPol (s.f.) tres indicadores que se usan en su herramienta al igual que la fórmula matemática de la cual tienen la patente.

En cuanto los datos, se informa de manera breve en la página web los datos que se utilizarán, en caso de que se contraten sus servicios. Según dicha información, el algoritmo necesitará únicamente datos de los delitos actuales e históricos. Además, no se hace mención alguna al origen de los datos de entrenamiento del sistema de IA.

Sobre la información referente al tipo de aprendizaje**,** al igual que se ha mencionado para la variable trazabilidad, se es poco claro en referencia a esta variable, mencionándose únicamente el uso de técnicas de ML.

También se observan claras faltas respecto la variable privacidad, sobre todo porque no tienen en ningún lugar nada respecto leyes. En RUBEL, CASTRO y PHAM (2020), tampoco se mencionan concretamente leyes específicas, pero se hace hincapié en la clara vulneración de derechos y libertades que se vulneran haciendo uso de este sistema de IA.

Por último, en referencia a los derechos fundamentales**,** se repite la vulneración a la protección de datos de carácter personal (art. 8 Carta de los Derechos Fundamentales de la Unión Europea) añadiéndose en este caso la posibilidad de existir sesgos en los datos de entrenamiento del sistema de IA, y, por lo tanto, vulnerar el derecho a la no discriminación (art. 21 Carta de los Derechos Fundamentales de la Unión Europea) y, por extensión, el derecho a la igualdad (art. 20 Carta de los Derechos Fundamentales de la Unión Europea). De hecho, este sistema de IA cesó su implementación a raíz de LUM e ISAAC (2016), que encontraron que el algoritmo usado en este sistema tenía sesgos en contra de los barrios más pobres y las minorías.

5.2.Ámbito delincuencial

En cuanto esta parte, cabe destacar que no hay aún ninguna sentencia en España dónde se expongan casos dónde ha existido uso de sistemas de IA para llevar a cabo conductas delictivas, y, por lo tanto, se expondrán casos hipotéticos haciendo uso de sistemas de IA conocidos popularmente. En consecuencia, los tres hipotéticos casos expuestos para la realización de dicha tabla serán, el primero, el uso de la herramienta ChatGPT para la realización de *phishing*; el segundo, la creación de imágenes sexuales falsas de personas sin su consentimiento; y, por último, la asistencia de un sistema de IA para programar *malware* con el objetivo de infectar un sistema de IA, con tal de alterar su procedimiento lógico. Pese a esta hipotetización de casos, esta tabla sigue cumpliendo su función en cuanto especifica características de los sistemas de IA expuestos.

Tabla 3. Sistemas de IA que se podrían usar en el ámbito de la criminalidad.

Sistemas de IA	Trazabilidad	Datos	Aprendizaje	Privacidad	Derechos	Uso
ChatGPT	Media	No	NLP+Transf.	Baja	Art. 8, 17	Réplica
DeepFaceLab	Alta	Sí	GAN+ML	Baja	Art. 1, 3, 7, 24 y 25	Contenido
GitHub Copilot	Baja	No	NLP*+ML	Baja	Art 17	Puro*

Fuente: Elaboración propia a partir de la información extraída de las fuentes citadas en sus apartados correspondientes.

5.2.1. *Phishing* mediante ChatGPT

Pese a que el ChatGPT es también un servicio ofrecido por una empresa privada, si existe más información sobre el mismo debido a su actual repercusión mediática, que ha obligado a la empresa a cambiar en referencia a alguna de las

variables expuestas en la tabla, por lo tanto, se usarán los datos extraídos de OpenAI (2022), para observar cuanto se ha mejorado.

Empezando por la primera de las variables, la trazabilidad, categorizada como ‘media’, ya que, a pesar de ser tarea complicada, el seguir de inicio a fin las elaboraciones de un sistema de IA, sobre todo a mayor complejidad de procesamiento, sí se dispone en su blog de una pequeña ilustración que muestra el procedimiento de los datos.

En referencia **a los datos**, si bien es cierto que si se conoce cierta información sobre hasta cuando abarcan los datos de entrenamiento del ChatGPT, concretamente hasta 2021, sigue teniendo lagunas en cuando a especificar con precisión las fuentes utilizadas, además de ocultar en cierta manera que tras su uso, el sistema se sigue alimentando con los datos proporcionados por los usuarios.

El tipo de aprendizaje como se ejemplificó en el marco teórico es el NLP combinado con una tecnología llamada ‘Transformers’, que a su vez es la combinación de otros tipos de aprendizaje por DL, concretamente, el AE y el RNN.

En cuanto la variable privacidad, esta ha sido el tema discutido mediáticamente tras la Garante (2023), pese a que ciertamente el ChatGPT incumple varios artículos del RGPD, han actualizado ligeramente sus bases normativas, añadiendo alguna opción de seguridad en relación con la protección de datos (OpenAI, 2023), sin embargo, siguen sin ser suficientes.

Como en todos los sistemas de IA, los derechos fundamentales que se verán afectados de manera primordial será el de la protección de datos de carácter personal (art. 8 Carta de los Derechos Fundamentales de la Unión Europea) y el derecho a la propiedad intelectual (art. 17.2 Carta de los Derechos Fundamentales de la Unión Europea). Además, como mención especial, este sistema de IA vulnera los art. 5, 6, 8, 13 y 25 RGPD (Garante, 2023).

Por último, la variable de tipo de uso del ‘sistema de IA’**,** estamos claramente en un uso del tipo réplica. En este caso, el ‘sistema de IA’ complementa los medios comisivos del comportamiento delictivo del *phishing*, que consiste en intentar

engañar a los usuarios a través del correo electrónico para que proporcionen al información valiosa al atacante (Myers, 2006; citado en Holt, 2019). En este contexto, la utilización de ‘sistemas de IA’ añade complejidad y credibilidad a sus acciones.

5.2.2. Creación de *DeepFakes*

Son muchos los sistemas de IA que permiten la creación, o más bien, modificación de videos e imágenes usando otros como base para crear contenido inédito. De base, es un tipo de herramienta que podría llegar a ser útil, pero como muchas otras herramientas, los cibercriminales también empiezan a usarlas para sus propios fines. DeepFaceLab ha sido el sistema de IA elegido, ya que se analiza en Petrov *et al.* (2021). Además, cabe destacar que es una herramienta que permite el entrenamiento autónomo por parte de los usuarios.

La trazabilidad está bien descrita en Petrov *et al.* (2021), en cuanto se describen los procedimientos del modelo de IA, como se analizan los datos y las fases por las que pasan los mismos.

En cuanto el origen de los datos en su entrenamiento, este sistema de IA tiene la ventaja, y a su vez la desventaja, de ser un sistema de IA de código abierto, por lo tanto, los datos de entrenamiento serán los que cada usuario quiera utilizar, se dispondrán, por lo tanto, de los datos base que se quieren editar y los datos con los que se editará el contenido primario.

Respecto el tipo de aprendizaje**,** utiliza el modelo GAN, unido a un modelo de aprendizaje supervisado, dónde el usuario será el que verifique si se está realizando o no la funcionalidad de manera correcta, y, en caso negativo, podrá continuar con el proceso de entrenamiento.

En cuanto, la característica de la privacidad**,** no se tienen en cuenta en ningún momento los reglamentos de protección de datos existentes, ni los derechos humanos, ni las posibles implicaciones que podrían tener sistemas de IA como estos.

Debido al elevado riesgo de este tipo concreto de sistema de IA, sobre todo en el caso hipotético elegido, los derechos fundamentales afectados son variados. Tenemos posibles vulneraciones del derecho a la dignidad humana (art. 1 Carta de los Derechos Fundamentales de la Unión Europea), debido la capacidad que tiene este sistema de IA de poner a cualquier ser humano en situaciones falsas que puedan comprometer su dignidad; vulneraciones del derecho a la integridad de las personas (art. 3 Carta de los Derechos Fundamentales de la Unión Europea), ya que un tercero se puede lucrar de la imagen sexual de otro volviéndolo objeto de lucro; vulneraciones en la esfera privada de las personas (art. 7 Carta de los Derechos Fundamentales de la Unión Europea), en cuanto se pueden mostrar aspectos íntimos de otro que pese a ser falsos no se pueden desmentir de fácil manera; y, finalmente, pude vulnerar los derechos de los menores (art. 24 Carta de los Derechos Fundamentales de la Unión Europea), en cuanto el programa no distingue entre menores o no, y, por lo tanto, se pueden usar sus imágenes para fines delictivos.

En referencia a la variable tipo de uso del sistema de IA, estamos hablando de un delito dónde el daño lo produce la distribución de un contenido, por lo tanto, se adapta el medio comisivo del delito a la utilización de sistemas de IA, pero no cambia el delito final.

5.2.3. Sistema de IA contra sistema de IA

A fin de disponer de un caso hipotético de cada ‘tipo de uso del sistema de IA’, el último caso hipotético planteado es el más complejo. Si bien es cierto que anteriormente se podría haber planteado el uso de sistemas como el ChatGPT que permitía, y sigue permitiendo en menor medida, corrección de código de programación, he decidido optar por un sistema de IA dedicado exclusivamente a la ayuda de programación. Debido a la complejidad del caso en sí mismo, la existencia de literatura al respecto es nula, de igual manera, para definir el sistema de IA se usarán los blogs oficiales de los creadores, en este caso GitHub (s.f.) y OpenAI (2021).

Primeramente, en cuanto a la categorización de la trazabilidad**,** viene dada por la nula explicación del procedimiento usado en ninguna de las fases del sistema de IA. Ya

que, únicamente se explica la base algorítmica en la página web de los desarrolladores y tampoco hay mucho más en OpenAI (2021), web de origen del algoritmo.

Respecto el origen de los datos, tienen un apartado dónde se describe con brevedad de dónde provienen los datos, aunque mayoritariamente GITHUB (s.f.) expresa que los datos vienen de código abierto existente en Internet e inclusive códigos guardados en la propia plataforma. Por lo tanto, los datos son poco definidos y podrían provenir de muchos sitios web diferentes con los riesgos que eso podría conllevar.

Lo relativo al tipo de aprendizaje tampoco es que sea expuesto de manera muy clara, pero si hacen mención del procesado de lenguaje natural, aunque en este caso concreto sería más bien lenguaje de programación. Además, como otros sistemas de IA que asisten la programación, es altamente probable que hayan pasado por una fase de aprendizaje por refuerzo, dónde la consecución del código objetivo les haría aprender.

Al igual que en anteriores ejemplos, la esfera de la privacidad sigue poco amparada por estos servicios privados, dónde únicamente se hace alusión a los servicios de privacidad que tienen como organización atendiendo al reglamento de su país, Estados Unidos, en este caso. A pesar de que se dedican ciertos apartados de la página web a responder preguntas respecto la privacidad, estás se quedan bastante escuetas y acaban enlazando con el reglamento anteriormente comentado.

En relación con los derechos fundamentales, principalmente, se vulneraría el derecho a la propiedad intelectual (art. 17.2 Carta de los Derechos Fundamentales de la Unión Europea), debido a la recogida de código de autores sin su debido consentimiento. También, debido a las características del caso concreto, se podría hacer mención a muchos otros derechos a vulnerar, pero estos dependerían del sistema de IA que se intenta atacar.

Por último, como ya he mencionado al principio de esta sección, el tipo de uso de la IA es puro, debido a la creación de un nuevo tipo de delito que ataca a los sistemas de IA, aunque también se podría considerar réplica si se califican penalmente los sistemas de IA como programario informático, pero de esto se hablará a continuación.

6. RESULTADOS

6.1.Análisis normativo comparado sobre las herramientas preventivas en uso

Después de examinar los sistemas de IA que se encuentran actualmente en uso, así como aquellos en desarrollo para un próximo despliegue en el ámbito de la seguridad, se procede a analizar su uso en consonancia con las normativas existentes, o en proceso de implementación, tanto en España como en la Unión Europea. En este sentido, el tercer sistema de IA, originalmente implementado en Estados Unidos, será evaluado considerando una hipotética aplicación en España, a fin de disponer de más ejemplos prácticos que comparar con la legislación presentada en este estudio.

6.1.1. Calificación de los sistemas de IA como de alto riesgo

Siguiendo lo expuesto en la futura 'Ley de Inteligencia Artificial' (2021), se confirmará, de manera preliminar, si los tres sistemas de IA analizados corresponden a aquellos definidos como 'de alto riesgo' en referencia al segundo apartado del art. 6 de esta misma ley. Para ello, se debe analizar si dichos 'sistemas de IA' están incluidos en alguno de los apartados disponibles en el Anexo III de esta misma propuesta.

Primeramente, el sistema VeriPol, encaja con la siguiente definición: "sistemas de IA destinados a utilizarse por parte de las autoridades encargadas de la aplicación de la ley para llevar a cabo evaluaciones de […] el riesgo para las potenciales víctimas de delitos" (Anexo III, apartado 6 a, 'Ley de Inteligencia Artificial', 2021). Por lo tanto, si se contemplaría como 'sistema de IA de alto riesgo' amparado por la normativa venidera.

Respecto el uso de ABIS, en ese mismo anexo en su primer apartado se mencionan: "sistemas de IA destinados a utilizarse en la identificación biométrica remota <<en tiempo real>> o <<en diferido>> de personas físicas" (Anexo III, apartado 1 a, 'Ley de Inteligencia Artificial', 2021). Por lo tanto, los sistemas automáticos de identificación biométrica también están comprendidos como de alto riesgo.

Por último, el sistema PredPol, también sería un 'sistema de IA de alto riesgo', ya que sus funcionalidades estarían comprendidas por varios apartados, primero:

“sistemas de IA destinados a utilizarse por parte de las autoridades encargadas de la aplicación de la ley para predecir la frecuencia de un infracción penal” y también los “destinados a utilizarse para llevar a cabo análisis sobre infracciones penales […], que permitan a las autoridades […] examinar grandes conjuntos de datos […] para detectar modelos desconocidos o descubrir relaciones ocultas en los datos.” (Anexo III, apartado 6 e y g, ‘Ley de Inteligencia Artificial’, 2021).

6.1.2. Evaluación del cumplimiento normativo de los sistemas de IA

Una vez observado que los sistemas de IA si estarían incluidos por la propuesta de reglamento, se analizará a continuación, si cumplen con los requisitos estipulados para dichos sistemas de alto riesgo de acuerdo con el capítulo 2 de la ‘Ley de Inteligencia Artificial’. En dicho capítulo, se imponen los siguientes requisitos de acuerdo con los sistemas de IA: un sistema de gestión de riesgos iterativo (art. 9); criterios de calidad para los datos utilizados en el entrenamiento (art. 10); documentación técnica (art. 11); capacidad de registros (art. 12); niveles de transparencia adecuados (art. 13); vigilancia humana (art. 14) y niveles adecuados de precisión, solidez y ciberseguridad (art. 15).

6.1.2.1. VeriPol el sistema de IA en uso

En la línea del primer requisito (art. 9 ‘Ley de Inteligencia Artificial’, 2021), es fundamental la implementación de un sistema de gestión de la seguridad de la información (SGSI) en las empresas que manejen información, que a su vez contemplará la gestión de los riesgos de manera iterativa, incluyendo el SGSI de la organización en su totalidad y los SGSI singulares de aquellos proyectos que incluyan manejo y gestión de datos conforme el RGPD (Codolà, s.f.). Por ende, este SGSI deberá de implementarse por parte de la organización que lleve a cabo dichos proyectos.

Actualmente, el sistema de IA en uso VeriPol, lo lleva a cabo la Policía Nacional, sin embargo, no hay indicios públicos de que se lleven a cabo estos sistemas de gestión ya que carecen de certificación oficial de la norma ISO/IEC 27001, la norma internacional que permite acreditar si una organización está siguiendo los requisitos técnicos para los SGSI (INCIBE, s.f.).

En cuanto a los criterios de calidad de sus datos, el art. 10 nos expone prácticas adecuadas de gobernanza para el conjunto de datos utilizado, elegir un correcto diseño, poner foco en una buena recopilación de datos, evaluar de manera previa la disponibilidad de los datos, entre otros, que se desconocen con la información pública disponible. Sin embargo, también destaca la necesidad de realizar exámenes previos para atender a posibles sesgos en los datos (art.10.2 f 'Ley de Inteligencia Artificial', 2021), y, como se expuso anteriormente este sistema tiene en sus datos un gran sesgo lingüístico (Eticas Foundation, 2021).

Como se ha observado anteriormente, el sistema VeriPol, sí dispone de documentación técnica en Quijano, Liberatore, Camacho y Camacho (2018), sin embargo, esta no cumple con la totalidad de lo estipulado en el art. 11 de la propuesta, en cuanto no incluye una descripción general del *software* que utiliza el sistema, las instrucciones de uso para el usuario, las descripciones físicas del soporte del sistema, descripción detallada del sistema de gestión de riesgos y otros documentos normativos disponibles en el Anexo IV de la 'Ley de Inteligencia Artificial' (2021).

Respecto la capacidad de generar registros automatizados explicada en el art. 12 de la propuesta, tampoco existe información publicada, pese a esto, es probable que si existan ya que los Cuerpos y Fuerzas de seguridad siguen protocolos de almacenamiento de registros para otro tipo de actividades policiales.

En cuanto a la transparencia y comunicación de información a los usuarios del art. 13 se desconoce si existen manuales internos para todos los trabajadores de la policía que garantice un adecuado entendimiento por su parte de cómo funciona dicho 'sistema de IA', se observa en Quijano, Liberatore, Camacho y Camacho (2018) que para el estudio piloto si se dio a los agentes locales una pequeña formación, a pesar de esto, en Pita (2020) se alerta de que existen comisarias dónde no se hace uso de esta herramienta debido la falta de formación al personal.

Siguiendo lo expuesto en el art. 14 de la propuesta, los 'sistemas de IA de alto riesgo' deberán de diseñarse para que puedan ser vigilados por personas físicas, sin embargo, en Quijano, Liberatore, Camacho y Camacho (2018), no hace mucha

incidencia en esta capacidad, únicamente se menciona la capacidad de automatización y agilización del proceso de recolección de denuncias.

Por último, en cuanto los últimos requisitos del art. 15 de la 'Ley de Inteligencia Artificial', se hace referencia en Quijano, Liberatore, Camacho y Camacho (2018), los niveles de precisión esperados y los parámetros utilizados para obtener dichos niveles, sin embargo, no se hace referencia alguna en la documentación oficial o en páginas de la policía nacional sobre aspectos de ciberseguridad.

Se observa por tanto, que el sistema VeriPol, deberá de ser actualizado antes de la implementación legislativa, o en su defecto, deberá ser cancelado o detenido si no se cumplen todos los requisitos establecidos.

6.1.2.2.Sistemas de IA en desuso o en vías de desarrollo

Se realizará un análisis conjunto de los otros 'sistemas de IA', considerando su estado preliminar o cancelado. Además, se mencionarán únicamente los requisitos de los que se dispone información debido a la falta de información actualmente disponible sobre los mismos, además de poderse extrapolar mucha de la información del anterior análisis, al ser estos 'sistemas de IA' aplicaciones para los Cuerpos y Fuerzas de seguridad.

Empezando por el sistema de gestión de riesgos del art. 9, se puede observar en Comisión Europea (2021), como la organización responsable de la implementación de ABIS, será el Ministerio del Interior, en conjunto con la policía, que, de nuevo, ninguno de estos dos organismos dispone de una certificación pública de ISO/IEC 27001, y, por tanto, se desconoce la práctica de SGSI en su contexto. Para PredPol, se puede extrapolar que serán esos organismos los que lo llevarían a cabo, por ende, tenemos el mismo problema.

Sobre la gobernanza de los datos del art. 10, el sistema PredPol carecía de la misma, al componerse en esencia por datos altamente sesgados (Lum e Isaac, 2016). En cuanto a ABIS, se conocen los datos de entrenamiento que se utilizarán, pero no el diseño, respecto su validación y prueba, por lo tanto, ninguno de estos cumplía o cumple actualmente este requisito.

En relación con la documentación técnica expuesta en el art. 11, ABIS no tiene aún una documentación técnica implementada de manera pública y formal debido a su actual desarrollo. El sistema PredPol, no disponía de documentación técnica adecuada, por ello se llevó a cabo la investigación de Lum e Isaac (2016).

En lo que respecta a los demás requisitos establecidos en los artículos 12, 13, 14 y 15 de la 'Ley de Inteligencia Artificial', no se dispone de información pública que nos brinde indicios sobre su cumplimiento.

En base a lo expuesto, será necesario considerar los requisitos mencionados tanto para la futura implementación de ABUS como para una eventual reedición de la herramienta PredPol, ya que se requerirá su cumplimiento obligatorio al tratarse de 'sistemas de IA de alto riesgo'.

6.2.La necesidad de una actualización legislativa para los delitos emergentes

A fin de resolver la cuestión planteada, se analizarán los delitos establecidos en el Código Penal español, que puedan llegar a implicar sistemas de IA. Cabe destacar que la mayoría de las implicaciones serán hipotéticas, ya que no existe actualmente jurisprudencia, a nivel nacional ni europeo, que haya valorado casos dónde se haya mediado el uso de sistemas de IA para realizar conductas delictivas.

Por ello, se analizará dónde encajarían los casos expuestos en la Tabla 3 de acuerdo con el Código Penal, sentencias judiciales, para justificar los tipos penales, y otras tipologías que pueden verse afectadas por el uso de sistemas de IA, a fin de vislumbrar la necesidad de añadir nueva legislación para dichos comportamientos delictivos.

El primer caso hipotético planteado ha sido el uso de sistemas de IA de procesamiento de lenguaje natural para mejorar las técnicas de *phishing* y/o ingeniería social. Como se expuso de manera anterior, el sistema de IA en este caso se utiliza como complemento del comportamiento 'tradicional' de *phishing*, cuyo tipo penal depende del fin último de la acción. Si se busca conseguir una transferencia no consentida, entraría en la estafa informática del art. 249 CP (SAP M 2642/2023, SAP IB 857/2023 y SAP Z 84/2023), en cambio, si el objetivo es acceder a sistemas de información o que

otros accedan, mediante las credenciales obtenidas, se entrará en los art. 197 bis y 197 ter CP respectivamente (SAP SO 138/2022 y SAP O 758/2022).

Además de esto, también es importante plantear los retos que supone este tipo de comportamiento en referencia a la responsabilidad de los sujetos implicados. Ya que es muy difícil imputar responsabilidad penal debido a múltiples factores, muchos *phishers* no utilizan los datos de la víctima, sino que los venden a terceros; la capacidad de rastreo de dichos comportamientos es muy reducida; la jurisdicción en casos de delitos transfronterizos suele funcionar más despacio y que las víctimas de estos tipos de fraude online suelen ser doblemente victimizadas y, por lo tanto, no tienden a denunciar los hechos sufridos (Miró, 2012; Button y Cross, 2017).

El segundo caso expuesto tiene relación con la creación de 'Deep Fakes', ya sean contenidos sexuales o no. Como ya se comentó en su apartado correspondiente, este tipo de comportamiento delictivo afecta a una multitud de derechos de las personas. Primeramente, puede generar afectaciones graves al derecho a la integridad moral, y, por lo tanto, entraría en el tipo penal del art. 173 CP. Además de esta calificación, también podría entrar este hipotético caso en un delito de injuria (art. 208 CP). En cambio, si la víctima de esta falsificación de imágenes fuera un menor de edad, podríamos hablar de la vulneración a los derechos de los menores (art. 24, Carta de los Derechos Fundamentales de la Unión Europea), por ende, si el contenido falsificado fuera de carácter sexual podría condenarse por un delito de explotación sexual a menores basado en imágenes expuesto en Agustina, Montiel y Gámez-Guadix (2020) correspondiente al art. 189 b) CP.

También será importante vislumbrar posibles complicaciones respecto los delitos contra la intimidad, ya que la divulgación de dichas imágenes falseadas por 'sistemas de IA' menoscabarían gravemente la intimidad personal de la víctima, independientemente de que sean reales o no, sobre todo si dichas imágenes no pueden ser contrastadas por la misma. Sin embargo, siguiendo lo expuesto en el art. 197.7 CP, la difusión de las imágenes o grabaciones se tienen que obtener con anuencia de la víctima, por ende, no podrían entrar en este tipo penal concreto. En este último caso concreto, podría crearse un nuevo supuesto que incluya la generación de imágenes o grabaciones que vulneren

gravemente la intimidad personal de las personas, sin necesidad de que las mismas fueran reales.

En base a los casos expuestos, la doctrina podría considerar la creación de una tipología delictiva específica para los 'Deep Fakes'. Esta nueva categoría abarcaría, por un lado, aquellos contenidos no sexuales que causen afectaciones en el honor y dignidad de las víctimas, y, por otro, los contenidos que afecten a su dignidad sexual, siendo este último considerado como una forma agravada del primero.

El último de los casos expuestos en el apartado anterior hace referencia al uso de sistemas de IA que ayuden a los cibercriminales a comprometer otros sistemas de IA, para poder interferir en sus procesamientos lógicos modificando sus resultados. Si bien es cierto que se calificó como tipo de uso puro este comportamiento, y, por lo tanto, se podrían añadir tipos penales nuevos al Código Penal que se fundamenten en dichos ataques a sistemas de IA. Será decisión del legislador añadir esta tipología o incluir este tipo de casos en los delitos de daños informáticos (art. 264 CP), debido a que la amplitud de su definición podría llegar a incluir sistemas de IA.

Aunque no se haya ejemplificado anteriormente, es importante mencionar los delitos relacionados con la propiedad intelectual de los datos utilizados para el entrenamiento de los sistemas de IA. Durante el estudio, se ha observado que son pocos los sistemas de IA que cuentan con una trazabilidad adecuada debido a su propia naturaleza. No obstante, existen algunos sistemas de IA que sí definen de manera clara el origen de los datos utilizados en su entrenamiento y funcionamiento, mientras que otros tampoco lo especifican.

En este sentido, será necesario considerar la posibilidad de incluir en los delitos relativos a la propiedad intelectual, mediante la actualización del art. 270 del Código Penal, especificaciones sobre los casos en los que el entrenamiento de sistemas de IA conlleve vulneraciones de la propiedad intelectual.

También se podría hablar de la responsabilidad penal de los proveedores de dichos 'sistemas de IA'. En este sentido, VALLS (2022), concluye que esta deberá de analizarse según la fase de desarrollo en la que se encuentre el 'sistema de IA' en

particular. En este hipotético caso planteado, la doctrina mayoritaria podría plantear la responsabilidad por imprudencia para los desarrolladores (Valls, 2022).

7. CONCLUSIONES

A modo de cierre, se procederá a exponer las conclusiones finales del trabajo, la discusión de las hipótesis planteadas así como aquellas limitaciones encontradas a lo largo de la investigación.

I. Se ha evidenciado que a pesar de todos los avances tecnológicos que han implicado los sistemas de IA, con sus consiguientes avances sociales, la legislación europea aún no está preparada para ellos. Se plantean buenas soluciones incorrectamente aplicadas. Es inteligente pensar en una distinción entre aquellos sistemas de IA que puedan poner en riesgo los derechos fundamentales más cruciales, como la vida y la libertad, y otros que no. Sin embargo, resulta preocupante que, a pesar de poner atención en la importancia de los datos, en cuanto sirven para capacitar estas novedosas máquinas inteligentes, dejan de lado su protección en aras del avance tecnológico, y por tanto económico de la Unión Europea.

II. Es bien sabido que el derecho suele ir rezagado respecto los cambios sociales que ocurren, y, por consiguiente, la investigación, el mercado y la utilización de sistemas disruptivos siempre aparecerá antes que la legislación. No obstante, debería de ser el derecho el que ponga fin a las malas praxis o el que directamente no permita el uso de herramientas aún no legisladas por el bienestar de la sociedad. Permitiendo salvaguardar la integridad y la equidad en el uso de sistemas disruptivos, evitando abusos y situaciones riesgosas para la sociedad en su conjunto.

En relación con las hipótesis del estudio, se ha podido obtener la siguiente información:

I. Tal y como se hipotetizó, la IA se utiliza de manera complementaria para la comisión de delitos, ya que hasta el momento, no ha planteado nuevos paradigmas delictivos tal y como si hizo la aparición de Internet.

II. Respecto la segunda hipótesis, se han ejemplificado y expuesto algunos casos dónde la aplicación de las herramientas para la investigación criminal no tuvo en cuenta posibles sesgos en la fase de entrenamiento, ya sean de carácter económico o intelectuales.

III. En cuanto la tercera hipótesis, si bien es cierto que la utilización de sistemas de IA por parte del sector de la seguridad plantea implicaciones en relación con los derechos fundamentales, al igual que ocurre con su uso por parte de ciberdelincuentes, se prevé que una vez la normativa esté vigente, dichas vulneraciones estarán sujetas a cuestiones de proporcionalidad en su uso. Por lo tanto, incurrirán en menor riesgo que el uso realizado por los cibercriminales.

IV. Finalmente, se ha constatado que los sistemas de IA no deberían de conllevar una necesaria reforma del Código Penal a gran escala, a excepción de la posible inclusión de los 'Deep Fakes' (fenómeno que no va intrínsecamente ligado a las IA). No obstante, es bien sabido que los sistemas de IA evolucionan de manera rápida y constante, desarrollando nuevas capacidades y funcionalidad, por lo tanto, en un futuro no tan lejano, es probable que estos avances tecnológicos requieran una reforma legislativa que se adapte a las demandas y desafíos planteados por las propias IA.

Además, durante la realización del estudio, se identificaron diversas limitaciones que resultan relevantes de destacar a fin de proporcionar una comprensión más completa de la revisión y los resultados:

I. Como se ha podido observar durante la investigación, las contemplaciones legales sobre IA se basan principalmente en una propuesta, y, por lo tanto, el estado final de las mismas puede llegar a influir en el entendimiento del estudio.

II. Además, la naturaleza disruptiva del tema ha dificultado la búsqueda de usos reales bien documentados en ambos sectores de interés para el trabajo. Un ejemplo de esto es la falta de sentencias, en CENDOJ, dónde se haga un buen uso de la terminología 'Inteligencia Artificial', ya que solo una de quince cumple con este criterio (AAP B 1448/2021). Asimismo, la ausencia de legislación vigente y la percepción de los

sistemas de IA como productos privados también han dificultado la obtención de datos para la parte técnica del estudio.

Como reflexión final, es fundamental destacar que los sistemas de IA actualmente operan sin una regulación efectiva. Será en el momento en que se establezcan las leyes pertinentes cuando los análisis sobre estos sistemas podrán hacerse de manera escrupulosa para determinar si realmente su uso es proporcional y si de verdad cumplen con los principios establecidos en el LIBRO BLANCO SOBRE LA INTELIGENCIA ARTIFICIAL (2020), es decir: “que la inteligencia artificial se debe asentar a nuestros valores y derechos fundamentales, como la dignidad humana y la protección de la privacidad”.

8. BIBLIOGRAFÍA

Agustina, J, Montiel, I. y Gámez-Guadix, M. (2020). *Cibercriminología y victimización online.* Editorial Síntesis.

Alonso, C. (2021). Acerca de la inteligencia artificial en el ámbito penal: especial referencia a la actividad de las fuerzas y cuerpos de seguridad. *Ius et Scientia, 7* (1), 25-36. https://doi.org/10.12795/IETSCIENTIA.2021.i01.03

Alonso, A. y Bolón, V. (2020). Inteligencia artificial, algoritmos y derecho. Una introducción. *Universitat Oberta de Catalunya.*

Bello, M. y Griffiths, M. (2021). Routine activity theory and cybercrime investigation in Nigeria: how capable are law enforcement agencies?. *Rethinking Cybercrime: Critical Debates*, 213-235. https://doi.org/10.1007/978-3-030-55841-3_11

Button, M., y Cross, C. (2017). Cyber Frauds, Scams and their Victims (1st ed.). Routledge. https://doi.org/10.4324/9781315679877

Carrillo, J. J. (2020). *Factor Humano: La Teoría de las Actividades Cotidianas en la Ciberseguridad.* https://www.academia.edu/download/65051818/Factor_Humano_TAC_en_Ciberseguridad.pdf

Codolà, S. (s.f.). *Seguridad y auditoría de la información*. Universitat Oberta de Catalunya.

Comisión Europea. (2021). *Summary Report of the project "Towards the European Level Exchange of Facial Images".* Telefi Project. https://www.telefi-project.eu/sites/default/files/TELEFI_SummaryReport.pdf

Cook, S., Giommoni, L., Trajtenberg, N., Levi, M. y Williams, M. (2023). Fear of economic cybercrime across Europe: A multilevel application of Routine Activity Theory. *The British Journal of Criminology, 63*(2), 384-406. https://doi.org/10.1093/bjc/azac021

Delgado, M. (1996-7). *La Inteligencia Artificial: Realidad de un mito moderno*. Universidad de Granada. https://digibug.ugr.es/bitstream/handle/10481/1270/16912512.pdf?sequence=1

España. Ley Orgánica 10/1995, de 23 de noviembre, del Código Penal. (BOE, núm. 281, 24-11-1995, pág. 33987-34058).

Eticas Foundation (2021). *Observatory of Algorithms with Social Impact, OASI* https://eticasfoundation.org/oasi/

Garante per la protezione dei dati personali. (2023). *Provvedimiento del 30 marzo 2023.* https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870832

GitHub (s.f.). *GitHub Copilot.* https://github.com/features/copilot

Gobierno de España (s.f.). *Dispositivos de control telemático de medidas y penas de alejamiento.* https://violenciagenero.igualdad.gob.es/informacionUtil/recursos/dispositivosControlTelematico/home.htm

González, V. (2022, diciembre 9). ¿Cómo será regulada la tecnología de reconocimiento facial que utilizará la Policía Nacional y la Guardia Civil?. *ConfiLegal.*

https://confilegal.com/20221209-como-sera-regulada-la-tecnologia-de-reconocimiento-facial-que-utilizara-la-policia-nacional-y-la-guardia-civil/

Grabosky, P. N. (2001). Virtual criminality: Old wine in new bottles?. *Social & Legal Studies, 10*(2), 243-249. https://doi.org/10.1177/a017405

Gupta, R., Srivastava, D., Sahu, M., Tiwari, S., Ambasta, R. y Kumar, P. (2021). Artificial intelligence to deep learning: machine intelligence approach for drug discovery. *Molecular diversity*, 25, 1315-1360. https://doi.org/10.1007/s11030-021-10217-3

Haenlein, M. y Kaplan, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. *California Management Review, 61*(4), 5–14. https://doi.org/10.1177/0008125619864925

Hawdon, J., Parti, K. y Dearden, T. (2020). Cybercrime in America amid COVID-19: The initial results from a natural experiment. *American Journal of Criminal Justice, 45*(4), 546-562. https://doi.org/10.1007/s12103-020-09534-4

Holt, T, Bossler, A y Seigfried-Spellar, K. (2018). Cybercrime and criminological theories. *Cybercrime and digital forensics: an introduction*. (pg. 439-490).

Holt, T. (2019). *El carding y el robo de datos*. Universitat Oberta de Catalunya.

Holt, T., Leukfeldt, R. y van de Weijer, S. (2020). An examination of motivation and routine activity theory to account for cyberattacks against Dutch web sites. *Criminal Justice and Behavior, 47*(4), 487-505. https://doi.org/10.1177/0093854819900322

Holt, T. y Steinmetz, K. (2021). Examining the role of power-control theory and self-control to account for computer hacking. *Crime & Delinquency, 67*(10), 1491-1512. https://doi.org/10.1177/0011128720981892

INCIBE (2019, octubre 10). *¿Conoces la nueva norma para la gestión de la privacidad?*. https://www.incibe.es/empresas/blog/conoces-nueva-norma-gestion-privacidad

Kigerl, A. (2021). Routine activity theory and malware, fraud, and spam at the national level. *Crime Law Soc Change*, 76, 109–130. https://doi.org/10.1007/s10611-021-09957-y

Klauser, F. (2021). Police Drones and the Air: Towards a Volumetric Geopolitics of Security. *Swiss Political Science Review, 1*(27), p. 158-169. https://doi.org/10.1111/spsr.12431

Luknar, I. (2022). Social control theory and cybercrime. *National Interest, 41*(1), 147-159. https://doi.org/10.22182/ni.4112022.7

Lum, K. y Isaac, W. (2016). To Predict and Serve?, *Significance*, *13*(5), p. 14–19, https://doi.org/10.1111/j.1740-9713.2016.00960.x

Miró, L. F. (2012). *El cibercrimen: Fenomenología y criminología de la delincuencia en el ciberespacio*. Marcial Pons Ediciones Jurídicas y Sociales.

Ministerio del Interior (2023, abril 4). *Registro de actividades de tratamiento del Ministerio del Interior.* https://www.interior.gob.es/opencms/pdf/servicios-al-ciudadano/participacion-ciudadana/proteccion-de-datos-de-caracter-personal/tutela-de-los-derechos/Registro_de_Actividades_de_Tratamiento_del_Ministerio_del_Interior.pdf

Moise, A. (2020). Cyber-criminology a new field of scientific research and criminological investigation. *Journal of Law and Administrative Sciences, 14*(14), 121-126. https://heinonline.org/HOL/LandingPage?handle=hein.journals/jladsc14&div=16&id=&page=

Navarro, J. y Marcum, C. (2020). Deviant Instruction: The Applicability of Social Learning Theory to Understanding Cybercrime. *The Palgrave Handbook of International Cybercrime and Cyberdeviance*, 527-545. https://doi.org/10.1007/978-3-319-78440-3_18

OpenAI. (2022, noviembre 30). Introducing ChatGPT. *OpenAI.* https://openai.com/blog/chatgpt

OpenAI. (2023). Data Controls FAQ. *OpenAI.* https://help.openai.com/en/articles/7730893-data-controls-faq

Ortega, D., Kanjo, E., Anwar, A., Johnson, S. y Lucy, D. (2021). *The rise of technology in crime prevention: Opportunities, challenges and practitioners' perspectives.* https://arxiv.org/abs/2102.04204

Partin, R., Meldrum, R., Lehmann, P., Back, S. y Trucco, E. (2022). Low self-control and cybercrime victimization: An examination of indirect effects through risky online behavior. *Crime & Delinquency, 68*(13-14), 2476-2502. https://doi.org/10.1177/00111287211061728

Perov, I., Gao, D., Chervoniy, N., Liu, K., Marangonda, S., Umé, C., ... y Zhang, W. (2020). DeepFaceLab: Integrated, flexible and extensible face-swapping framework. *Cornell University*. https://doi.org/10.48550/arXiv.2005.05535

Pita, E. (2020, agosto 30). La comisaría de Vigo lleva dos años sin usar la <<máquina de la verdad>>. *La Voz de Galicia*. https://www.lavozdegalicia.es/noticia/vigo/vigo/2020/08/30/comisaria-vigo-lleva-dos-anos-usar-maquina-verdad/0003_202008V30C1991.htm

PredPol (s.f.). *PredPol.* https://www.predpol.com/

Quevedo, J. (2017). *Investigación y prueba del ciberdelito.* [Tesis doctoral, Universitat de Barcelona]. Tesis Doctorals en Xarxa (TDX). https://www.tdx.cat/handle/10803/665611

Quijano, L., Liberatore, F., Camacho, J. y Camacho, M. (2018). Applying automatic text-based detection of deceptive language to police reports: Extracting behavioral patterns from a multi-step classification model to understand how we lie to the police. *Knowledge-Based Systems,* 149, 155-168. https://doi.org/10.1016/j.knosys.2018.03.010

Quijano, L., Liberatore, F. y Camacho, M. (2019). Applications of data science in policing: VeriPol as an investigation support tool. *Special Issue 4 Eur. Police Sci. & Res. Bull.*p. 89-96. https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/elerb4000&section=13

Rai, A. (2019). Explainable AI: from black box to glass box. *Journal of the Academy of Marketing Science,* 48, 137-141. https://doi.org/10.1007/s11747-019-00710-5

Rubel, A., Castro, C. y Pham, A. (2021). Democratic Obligations and Technological Threats to Legitimacy: PredPol, Cambridge Analytica, and Internet Research Agency. *Cambridge University Press,* p. 163-183. https://philpapers.org/rec/RUBDOA-3

Sarker, I. (2021a). Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions. *SN Computer Science,* 2, 420. https://doi.org/10.1007/s42979-021-00815-1

Sarker, I. (2021b). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science,* 2, 160. https://doi.org/10.1007/s42979-021-00592

Sentencia de la Audiencia Provincial de Barcelona 1448/2021, del 15 de febrero de 2021.

Sentencia de la Audiencia Provincial de Oviedo 758/2022, del 18 de febrero de 2022.

Sentencia de la Audiencia Provincial de Soria 138/2022, del 4 de abril de 2022.

Sentencia de la Audiencia Provincial de Zaragoza 84/2023, del 26 de enero de 2023.

Sentencia de la Audiencia Provincial de Madrid 2642/2023, del 6 de marzo de 2023.

Sentencia de la Audiencia Provincial de Palma de Mallorca 857/2022, del 21 de marzo de 2023.

Shadmanfaat, S., Howell, J., Muniz, C., Cochran, J., Kabiri, S. y Fontaine, E. (2020). Cyberbullying perpetration: An empirical test of social learning theory in Iran.

Deviant Behavior, 41(3), 278-293. https://doi.org/10.1080/01639625.2019.1565513

Suh, J., Choe, J. y Park, J. (2020). A lifestyle-routine activity theory (LRAT) approach to cybercrime victimization: An empirical assessment of SNS lifestyle exposure activities. *Asia Pacific Journal of Information Systems, 30*(1), 53-71. https://www.earticle.net/Article/A371221

Thales. (s.f.). *Thales Cogent CABIS 7.0.* https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/biometrics/biometric-software/automated-biometric-identification-system/cabis-7-0

Thales. (2022, junio 16). La tecnología de Thales seleccionada para el nuevo sistema de entrada / salido del espacio Schengen en España. *Thales Press.* https://www.thalesgroup.com/es/el-mundo/group/press_release/tecnologia-thales-seleccionada-para-el-nuevo-sistema-entrada-salida

Turing, A. (1936). On computable numbers, with an application on the Entscheidungsproblem. *Journal of Math, 5*(58), 230-265. https://www.wolframscience.com/prizes/tm23/images/Turing.pdf

Unión Europea (2020, febrero 19). Libro Blanco sobre la Inteligencia Artificial – un enfoque orientado a la excelencia y la confianza. https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52020DC0065

Unión Europea (2020). Derechos de propiedad intelectual para el desarrollo de las tecnologías relativas a la inteligencia artificial. Resolución del Parlamento Europeo, de 20 de octubre de 2020, sobre los derechos de propiedad intelectual para el desarrollo de las tecnologías relativas a la inteligencia artificial. https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=OJ:JOC_2021_404_R_0007&from=EN

Unión Europea (2012, octubre 26). Carta de los Derechos Fundamentales de la Unión Europea. https://www.europarl.europa.eu/charter/pdf/text_es.pdf

Unión Europea (2021a). Propuesta de Reglamento del Parlamento Europeo y del Consejo, por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión. https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex:52021PC0206

Unión Europea (2021b). Anexos de la Propuesta de Reglamento del Parlamento Europeo y del Consejo, por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión. https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex:52021PC0206

Valls, J. (2022). Sobre la responsabilidad penal por la utilización de sistemas inteligentes. *Revista electrónica de Ciencia Penal y Criminología*, 24, p.1-35. http://criminet.ugr.es/recpc/24/recpc24-27.pdf

Vaswani, A., Shazeer, N., Parmar, N., Uskoreit, J., Jones, L., Gomez, A., Kaiser, Ł. y Polosukhin, I. (2017). Attention Is All You Need. *Advances in neural information processing systems, 30.* https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Vestri, G. (2023, mayo 19). Mucho ChatGPT, pocos “algoritmos verdes”. *Observatorio Sector Público e Inteligencia Artificial.* https://www.ospia.org/o-lab/mucho-chatgpt-pocos-algoritmos-verdes

Wang, P. (2019). On Defining Artificial Intelligence. *Journal of Artificial General Intelligence,10*(2) 1-37. https://doi.org/10.2478/jagi-2019-0002