Universidad Rey Juan Carlos | Facultad de Ciencias Jurídicas y Políticas

TRABAJO FIN DE GRADO

GRADO EN CIENCIA POLÍTICA

CURSO ACADÉMICO 2022/2023

CONVOCATORIA

INTELIGENCIA ARTIFICIAL Y ADMINISTRACIÓN PÚBLICA: Un estudio de caso sobre los riesgos para el ciudadano en la prestación de bienes y servicios públicos.

El caso de VeriPol

AUTORA: Del Castillo Gil, Elena

GRUPO: Triple grado en Filosofía, Ciencia Política y Economía

TUTOR: García-Vegas, Ricardo

En Madrid, a 27 de octubre de 2023

A mis compañeros. Ha sido un camino largo, pero llegamos al final.

A mis padres, por la infinita paciencia todos estos años, a mi madre por ser firme pilar de apoyo incondicional emocional y a mi padre por ser el revisor jefe de todos mis trabajos, también de los TFGs,

A Ricardo, porque hacer una tutoría de TFG con él siempre te llenaba de energía y te levantaba el ánimo. Y porque durante la pandemia intentaba levantarnos el ánimo preguntándonos qué tal estábamos y qué habíamos hecho durante el día.

Muchas gracias a todos.

ÍNDICE

RESUMEN

La Inteligencia Artificial (IA) cada vez está más presente en la sociedad y es una tecnología capaz de revolucionar la vida diaria de los ciudadanos y de sus instituciones públicas. La Administración Pública ha comenzado a implantar en la prestación de servicios públicos este tipo de herramientas. La IA puede ser una gran aliada a la hora de obtener una información precisa sobre lo que necesitan los ciudadanos y así, adecuar mejor la oferta de servicios públicos por parte de la Administración. Sin embargo, la implementación de esta tecnología puede acarrear una serie de riesgos. A través del estudio de la aplicación VeriPol de detección de denuncias falsas, se examinarán los posibles riesgos a los que el ciudadano se enfrenta si dicho servicio se presta con ayuda de la IA.

Palabras Clave: IA, administraciones públicas, servicio público, ciudadano, riesgos.

ABSTRACT

Artificial Intelligence (AI) is increasingly present now in society and is a technology capable of revolutionising the daily lives of citizens and their public institutions. Public administration has begun to implement this type of tool to provide public services. AI can be a great ally when it comes to obtaining accurate information about what citizens need and thus better adapt the public services offered by the Administration. However, the implementation of this technology may entail several risks. Through the case study of the VeriPol application for detecting false reports, we will examine the possible risks that citizens face if this service is provided with the help of AI.

Key Words: IA, public administration, public service, citizens, risks.

CAPITULO I. PROTOCOLO DE LA INVESTIGACIÓN

1.1. Planteamiento

Las Administraciones Públicas (AA.PP.) son uno de los pilares de funcionamiento de nuestros sistemas políticos actuales y no escapa de los diferentes cambios que acontecen en la realidad sociopolítica de los países (Weber, 1947). Por ello, es inevitable que las nuevas herramientas que se desarrollan en el ámbito científico-tecnológico influyan en la configuración de la organización y estructuración de las propias Administraciones, así como en el diseño, ejercicio y aplicación de las políticas y servicios públicos que llevan a cabo. Las nuevas tecnologías de la información pueden modificar los procesos que se dan en el ámbito del sector público, al igual que ocurre en el sector privado (Ramió, 2019). No se debe de olvidar que las instituciones de un país reflejan y se encuentran en estrecha relación, aunque a veces no lo parezca, con la sociedad y los cambios que en ella acontecen.

La tecnología de la Inteligencia Artificial (IA) se puede aplicar en diferentes tareas y actividades propias de las instituciones públicas, desde la ejecución de políticas públicas para una mejor implementación hasta la prestación de un servicio público (Comisión Europea, 2020). Su uso puede ser muy beneficioso para los ciudadanos y es por ello por lo que estas herramientas requieren de nuestra atención y se ha de analizar si verdaderamente su empleo es eficiente a la hora de prestar servicios públicos que cubran necesidades específicas de los ciudadanos. Por ello, se ha de plantear una investigación acerca de cómo la Administración Pública ha respondido ante tales avances tecnológicos y su integración en el desarrollo de las políticas y servicios públicos, especialmente cuando solo una pequeña porción (4%) de los estudios publicados estudian la IA en el sector público (Comisión Europea, 2020a).

Las nuevas tecnologías de la información transforman los procesos que se dan en el ámbito del sector público. Sin embargo, cabe preguntarse si la aplicación que tienen este tipo de herramientas, cuyo desarrollo está afianzado en el ámbito privado, puede trasladarse también al ámbito público. Otra de las cuestiones a plantearse es en qué servicios públicos podría implementarse dichas herramientas, puesto que los algoritmos en los que se basan estas tecnologías realizan desde análisis de datos hasta proponer diferentes vías de actuación ante una problemática [Oliver, 2020; Filgueiras, 2021]. Cierto es que según vaya ampliándose su implantación, se irá aprendiendo cuál es su aplicación más efectiva y eficiente, al igual que en qué ámbitos es en los que debe aplicarse (Criado, 2021). Pero, ello no exime de hacer una cierta reflexión previa.

En muchos casos, la introducción de estas nuevas tecnologías es vista como una oportunidad para llevar a cabo una reforma profunda de mejora de las instituciones públicas (Ramió, 2018). Arenilla (2021) señala tres ideas que no deben pasar desapercibidas. En primer lugar, se ha de tener presente la importancia de la innovación pública y el conocimiento para volver a unir las demandas de los ciudadanos y las acciones de la Administración. En segundo lugar, estas tecnologías parecen presentar la oportunidad de renovar las instituciones y evitar así su obsolescencia. Y, en tercer lugar, gracias a estas tecnologías se presenta la oportunidad de una efectiva satisfacción de las demandas ciudadanas por parte de la Administración. Todo ello supondría un cambio en los valores, creencias, principios y cultura de las organizaciones para adaptarse a este

nuevo contexto (Arenilla, 2021, p.14), lo que conllevaría, en definitiva, una transformación integral de la organización.

Las Administraciones han de estar conectadas con sus ciudadanos para poder satisfacer las demandas que estos les presentan (Filgueiras, 2021). La IA parece presentarse como una buena herramienta para salvar esa distancia entre ambos, sin embargo, cabe plantearse algunas cuestiones al respecto. Este tipo de tecnología es una *tabula rasa*, que necesita de un entrenamiento y una base de datos específica para poder funcionar. Dicho entrenamiento y base de datos, por norma general son proporcionados por seres humanos, pero cabe preguntarse hasta qué punto dicho entrenamiento e información representa a la sociedad del país a la que se dirige, contiene una serie de sesgos o su uso supone una amenaza (Oliver, 2020, p. 26).

En los últimos cinco años a nivel mundial ha aumentado exponencialmente el empleo de la IA en el sector privado y ello también se refleja en el sector público, al que puede reportar grandes beneficios, siempre que se respete el nivel de satisfacción y la calidad de gobernanza de los servicios públicos (Misuraca & van Noordt, 2020, p.11). Aunque haya habido un crecimiento exponencial del uso de estas tecnologías, no todos los países se han desarrollado por igual (Campos, 2019). Según datos de la Comisión Europea, en 2016 aproximadamente América del Norte fue la región que más invirtió en inteligencia artificial (12.100-18.600 millones de euros), seguida de Asia (6.500-9.700 millones de euros) y Europa (2.400-3.200 millones de euros) (Parlamento Europeo, 2022).

Además, el impacto económico que se espera por el uso de robots, vehículos autónomos y la automatización de conocimientos es de 6.500-12.000 millones de euros al año, por lo que los impactos que esta tecnología ha tenido y tendrá en los siguientes años son significativos. Asimismo, se ha de considerar el número de solicitud de patentes de IA, que ha aumentado en un 400% en la última década. El número de solicitudes en el período 1960-2019 fueron en EE. UU. de 1.863, en China de 1.085 y en la UE de 1.074 (Parlamento Europeo, 2022).

Viendo la potencialidad que podía suponer la implementación de la IA a nivel económico político y social, en junio de 2018 la Comisión Europea se sumó a la Comunicación "Inteligencia Artificial para Europa" que países miembros de la Unión Europea, Suiza y Noruega habían adoptado en abril de ese año (Misuraca & Van Noordt, 2020). Aunque, se pueden encontrar antecedentes de la preocupación por el uso de la Inteligencia Artificial en el año 2016, con el informe elaborado por el parlamento británico, *Robotics and artificial intelligence* (House of the Commons, 2016), en el que se mencionaban aspectos como las posibles implicaciones del uso de esta tecnología, las oportunidades que esta podría ofrecer, los problemas sociales y económicos que podía generar y las medidas que el Gobierno habría de adoptar en relación a lo anterior (Campos, 2019, p. 81).

La pandemia del COVID-19 ha acelerado la revolución digital en muchos ámbitos, ya que en muchos casos el único modo de relacionarse con la Administración fue de forma telemática (Fundación Telefónica, 2021). La pandemia ha acelerado el uso de este tipo de tecnologías, ya que ha reducido las cargas de trabajo y ha facilitado la realización de procesos en la organización de las AA.PP. y ha cambiado la forma en la que las organizaciones públicas se relacionan con los ciudadanos. Por ejemplo, un uso sencillo de inteligencia artificial son los *chatbots* de atención ciudadana que clasifican nuestras consultas y nos redireccionan a quien pueda resolverlas, ahorrando tiempo de

espera del ciudadano para resolver su duda y haciendo más eficiente la labor de la institución[1].

La IA ha supuesto una gran revolución tecnológica, que especialmente se ha desarrollado en los últimos años del siglo XX. La Inteligencia Artificial se ha convertido en una de las herramientas tecnológicas principales y ha comenzado a salir de los centros de investigación para aplicarse en diferentes ámbitos y no solo el científico (Oliver, 2020).

Proporcionar una definición o conceptualizar la IA es complicado, pero para facilitar nuestro análisis en el presente trabajo, tomaremos de referencia la definición que se establece en la normativa de la Unión Europea. En el Libro Blanco sobre la inteligencia artificial (Comisión Europea, 2020b, p. 20), se establece que conforme a lo establecido por la Comisión Europea (COM (2018) 237 final, p. 1):

> *«El término "inteligencia artificial" (IA) se aplica a los sistemas que manifiestan un comportamiento inteligente, pues son capaces de analizar su entorno y pasar a la acción —con cierto grado de autonomía— con el fin de alcanzar objetivos específicos. Los sistemas basados en la IA pueden consistir simplemente en un programa informático (p. ej. asistentes de voz, programas de análisis de imágenes, motores de búsqueda, sistemas de reconocimiento facial y de voz), pero la IA también puede estar incorporada en dispositivos de hardware (p. ej. robots avanzados, automóviles autónomos, drones o aplicaciones del internet de las cosas)».*

Quizás se comprenda mejor el funcionamiento de esta tecnología y qué es exactamente si se analizan sus aplicaciones. Si se sigue una de las máximas que suelen abanderar los ingenieros ("*learning by doing*"), se comprenderá mejor esta tecnología cuando se vean los diferentes avances específicos que se han llevado a cabo en el ámbito en el que se aplican. Es más, una de las maneras más eficaces de comprobar si este tipo de recursos funcionan o tienen fallos en su diseño es comprobar cómo evoluciona cuando se aplica al campo sobre el que tiene que trabajar.

Por ello, la motivación del presente trabajo es analizar qué supone la introducción de la IA en la prestación de servicios públicos, con una especial atención a los riesgos que la aplicación de estos sistemas puede tener en la ciudadanía. Para estudiar estas cuestiones se examinará el servicio público VeriPol, para cuya ejecución se necesita de la herramienta tecnológica de la Inteligencia Artificial.

1.2. Justificación

Se ha hecho referencia a que no se ha prestado suficiente atención a la relación entre Inteligencia Artificial y la Administración Pública, por ello es necesario establecer un análisis de la relación existente entre ambos elementos (Filgueiras, 2021). Es de sumo interés ver cómo la progresiva digitalización de las instituciones públicas y la

[1] Un ejemplo es los *chatbots* que se encuentran en la página de inicio de muchos sitios web, que ayudan a filtrar, redireccionar o resolver las dudas básicas de los usuarios. Reduce la carga de trabajo de los empleados encargados de responder las dudas, si el *chatbot* no es capaz de responder la duda del ciudadano, entonces el empleado le atiende.

implementación de la IA en la elaboración de políticas y estrategias públicas, así como en su ejecución o monitoreo, incide en la correcta prestación de servicios públicos a los ciudadanos. Para el estudio de estas problemáticas han de tenerse en cuenta una serie de aspectos que remarcan la importancia del análisis de la interrelación entre la IA y la prestación de los servicios públicos, teniendo siempre presente la centralidad que debe de tener el ciudadano.

En el ámbito teórico son de suma importancia una serie de puntos relevantes que ayudan a clarificar el porqué de la importancia de hablar sobre estas cuestiones. Cada vez está más presente la IA en el ámbito práctico, por lo que hemos de plantearnos qué entraña la relación que habrá de darse entre las instituciones y la IA. A lo que se añade, que puede que se haya dado un paso más en la electronificación de la administración, la smartificación[2] [Ramió, 2019; Arenilla, 2021].

Por una parte, se ha de considerar que estas cuestiones están cada vez más presentes tanto en el ámbito teórico como práctico. Los cambios que se están produciendo en los elementos externos de las instituciones, inciden también en los elementos internos de estas. Los cambios que se producen en el contexto externo a la organización pueden influir determinantemente en su forma de organizar y establecer una serie de pautas de actuación y de ejecución, tanto de las políticas públicas como de los servicios públicos objetivo. Por ejemplo, Manuel Arenilla (2021) y Carles Ramió (2019) ponen el foco en la importancia de adaptar las organizaciones públicas a las nuevas dinámicas sociales, de manera que pueda preservarse los valores democráticos y el bienestar de la sociedad. Sin embargo, ello va unido a un cambio en las lógicas de funcionamiento que hasta ahora se estaban llevando a cabo, y a una necesidad de establecer una precisión conceptual, ya que para poder saber cómo implementar las nuevas herramientas tecnológicas se habrá de precisar qué es cada nuevo elemento.

Por otra parte, se ha de considerar que la Administración ya no se encuentra en su fase de "electronificación", sino que ha dado un paso más y se encuentra en la fase de "smartificación" (Ramió, 2019). A pesar de que aún se esté implementando y asentando la electronificación de la Administración[3] y se haya avanzado bastante en ese terreno, se ha mostrado como poco a poco las instituciones han podido servirse de las nuevas tecnologías para agilizar sus procesos. Con la smartificación, se da un paso más en dicha electronificación, puesto que es la consecuencia lógica de la revolución digital acontecida en los últimos años, debido al desarrollo de la IA. La IA no solo permite seguir utilizando los recursos electrónicos anteriores, sino que los mejora y ayuda a incrementar la calidad, eficacia y adecuación de la prestación de los servicios públicos a las necesidades de los ciudadanos.

[2] Smartificación es un término que acuña Ramió y refiere al siguiente paso en la electronificación de las AA.PP. en la que se emplea internet y las aplicaciones derivadas de su uso, y la IA y las aplicaciones derivadas de su uso (Ramió, 2019). "Por *smartificar* en el ámbito público se entiende el uso global, intensivo y sostenible de las tecnologías de la información bajo el principio de servicio para la mejora de calidad de los ciudadanos" (Ramió, 2019, p.13) "Para la smartificación de la Administración pública deberíamos proponer la siguiente definición: 'Proceso para lograr mayor inteligencia institucional para gobernar las complejas redes públicas y privadas con el objetivo final de aportar valor público a las actividades administrativas y atender de manera proactiva las necesidades de la ciudadanía'." (Ramió, 2019, p.14)

[3] Véase Ley 11/2007, de 22 de junio, de Acceso Electrónico de los Ciudadanos a los Servicios Públicos, Ley 39/2015, de 1 de octubre, del Procedimiento Administrativo Común de las Administraciones Públicas y Ley 40/2015, de 1 de octubre, de Régimen Jurídico del Sector Público.

A ello se suma que se ha de tener en consideración, que las mejoras que trae consigo la IA se insertan en el marco de la innovación pública. El Manual de Oslo (2018) y el Manual de Frascati (2015) elaborados por la OCDE establecen las premisas básicas más importantes para tener en cuenta si queremos utilizar de forma responsable los nuevos avances tecnológicos que la innovación presenta para recoger, reportar y usar datos. El Manual de Oslo (2018) en su cuarta edición, entiende por innovación lo siguiente (OCDE, 2018, p.20):

> *"An innovation is a new or improved product or process (or combination thereof) that differs significantly from the unit's previous products or processes and that has been made available to potential users (product) or brought into use by the unit (process)."*

A pesar de que cuando se habla de innovación solo se piensa en empresas privadas como Google o Amazon, entre otras, la innovación también está presente en las entidades públicas. Los organismos públicos no dejan de estar comprometidos con la mejora de la sociedad y los servicios públicos gracias a las diferentes innovaciones que se produzcan. Además, han de generar el ambiente propicio para que se den estas innovaciones y que suponga un desarrollo social y económico (Corvalán, 2017). Todo ello teniendo presente que quizás la colaboración público-privada sea necesaria debido a que muchas de las nuevas actualizaciones de este tipo de tecnologías provienen del ámbito privado (Ramió, 2019).

Pero todos estos procesos no pueden realizarse sin un marco jurídico que regule las actuaciones del sector público y privado, y las características de la tecnología que se desarrolle. Las regulaciones y normativas bajo la que se regirán y se insertarán la mayoría de las problemáticas que surjan entre las AA.PP. y la IA es el derecho administrativo (Arenilla, 2021). Aunque no solo el derecho administrativo se abre a este cambio de tendencia en la sociedad, sino que cada vez hay una mayor apertura por parte del sistema normativo a incorporar las diferentes dinámicas que surgen en la sociedad, especialmente con esta tecnología. Y si no existen problemas en el ámbito privado para el acceso y utilización de la IA, tampoco debiera de haber problemas en el ámbito público (Campos, 2019, p.77), prueba de ello es la nueva normativa sobre IA de la UE que enfatiza estas cuestiones (Comisión Europea, 2020b) y recalca, además, la necesidad de la colaboración público-privada (AI HLEG, 2019).

Como se ha comentado, el empleo de la IA es un elemento clave en la innovación pública y en las nuevas formas de proceder de las AA.PP.. Las actividades que realice la IA en el ámbito público han de insertarse en un marco jurídico concreto que proporcione las garantías necesarias para el buen funcionamiento de la administración y la protección de los ciudadanos frente al mal uso de estas herramientas. Por ello, es necesario analizar cómo la IA se introduce en el proceso político y de funcionamiento de las administraciones, a la hora de tomar decisiones o gestionar y proteger los datos personales de los ciudadanos. Los cambios que produce y producirá la IA están ligados con cómo se concibe la administración digital y cómo esta ha evolucionado y evolucionará en los siguientes años (De la Sierra, 2020).

En el ámbito empírico, se ha de destacar una serie de hechos y circunstancias presentes en el contexto político, económico, social y tecnológico a nivel nacional e internacional que dotan de relevancia a esta cuestión. En los últimos años la IA ha adquirido una creciente popularidad. Si se observan las noticias publicadas sobre tecnología en los últimos meses, se verá que la IA ha estado muy presente con la

evolución de los nuevos *chatbots*. Es una tecnología que es muy eficaz a la hora de obtener y recopilar información y facilitar el análisis de datos necesario, por ejemplo, para la definición y ejecución de políticas y servicios públicos.

Dada la creciente popularidad y aumento de su uso en el ámbito privado, las instituciones de los países se han visto en la necesidad de legislar y crear un marco normativo acerca de estas nuevas herramientas tecnológicas. La Unión Europea ha desarrollado el Libro Blanco sobre la Inteligencia Artificial y la política europea "*Artificial Inteligence*". Actualmente hay un grupo de estudio en Bruselas que reúne a expertos sobre la materia – AI HLEG- para asesorar sobre la elaboración de dichas normativas (AI HLEG, 2019). En España, hay una estrategia a nivel nacional (Gobierno de España, 2020), que se encuentra inserta en la Agenda Española Digital 2016, enmarcada en el componente 16 del Plan de Recuperación, Transformación y Resiliencia [Gobierno de España, 2020; Gobierno de España, 2021].

Asimismo, se pueden encontrar varios ejemplos de lo que se ha mencionado anteriormente. Un ejemplo es el caso del Gobierno de Andorra. El Gobierno de Andorra adquirió los datos personales recogidos por las tres principales compañías telefónicas que operaban en el país y cedieron dichos datos al *Massachusetts Institutte of Technology* (MIT). Con esta cesión de datos, se buscaba adquirir una mayor capacidad de análisis de los datos de sus ciudadanos, mediante la aplicación de nuevos métodos de análisis que empleaban IA, para detectar las necesidades de los andorranos. Estos nuevos métodos de análisis permitieron tener una información más exacta y precisa sobre las necesidades de los ciudadanos y así, gracias a ello, mejorar los servicios públicos que el gobierno andorrano prestaba (Ramió, 2019).

Otros ejemplos, son los diferentes servicios públicos que se han implementado en los últimos años en la Unión Europea, en el contexto de la *AI Watch*, el *Servicio de Conocimiento de la Comisión Europea para la supervisión del desarrollo, adopción e impacto de la Inteligencia Artificial para Europa*, que proporciona un análisis del uso e impacto de la IA en los servicios públicos a nivel local y estatal (Comisión Europea, 2020a). Algunos de los servicios que se presentan son: Tengai, servicio público de Suiza que mediante la IA quiere mejorar la administración y el uso de recursos y mejora del cumplimiento de las subvenciones; SyRi, servicio público de los Países Bajos, que mejora las capacidades de inspección y bienestar social y la reducción del uso de los fondos públicos; y VeriPol, servicio público de España, que ayuda a la detección de informes falsos y detección de informes fraudulentos (Misuraca & Van Noordt, 2020).

1.3. Objetivos

Una vez consideradas algunas de las razones que hacen relevante el estudio de la relación entre la IA y la Administración, se han de especificar los objetivos que se persiguen con el presente estudio de la implementación de la IA en los servicios públicos.

Una de las principales cuestiones a examinar son los riesgos que tiene la aplicación de estas tecnologías en los servicios públicos. El ciudadano ha de ser el foco principal bajo el que se examinen dichos riesgos, puesto que, en definitiva, es el ciudadano el destinatario de los servicios públicos que presta la Administración y el porqué de la implementación de unas políticas públicas u otras.

El funcionamiento de estas tecnologías innovadoras afecta de forma directa al ciudadano, ya que puede ver amenazados sus derechos fundamentales. La mayoría de los sistemas de IA se basan en la recogida de datos, lo que en algunas ocasiones resultará ser recogida de datos personales de los ciudadanos, amenazando así su intimidad e integridad en caso de un hackeo al sistema de datos, por ejemplo. Asimismo, corren el riesgo de que la tecnología empleada sea sesgada y no considere a todos los ciudadanos por igual o no se adapte correctamente a la especificidad requerida por la situación.

Además, la inteligencia artificial es una tecnología capaz de producir cambios profundos en el comportamiento y en los patrones de funcionamiento de una sociedad. La IA produce no solo un cambio en la forma de organización de las instituciones tanto públicas como privadas, sino que puede suponer un cambio en las relaciones que se puedan dar de forma interna y externa entre las administraciones y los ciudadanos. A lo que se añade el peligro de que dichos cambios se den como consecuencia de un tratamiento de datos que provenga de una información que sea inconsistente, sesgada o parcial.

Para llevar a cabo dicho análisis se ha de definir una serie de conceptos y teorías clave que van a ser recurrentes a lo largo de toda la investigación y que harán de guía en el análisis. Como pueden ser, por ejemplo, los conceptos de “inteligencia artificial” o “gobernanza de datos”, entre otros. Una vez establecidos, así como definidos ciertos modelos ideales de organización de las instituciones públicas y la gestión de los datos, se procederá a la exposición y análisis del servicio público elegido VeriPol. Realizada la exposición y el análisis de la información adquirida, se obtendrán una serie de resultados que revelarán si resulta beneficiosa la relación entre la IA y la Administración Pública en la ejecución y prestación de servicios públicos a los ciudadanos.

Tabla 1. Tabla de objetivos

Objetivo principal	Objetivos específicos
Analizar los principales riesgos a los que se ve expuesto el ciudadano con la implementación de la IA en los servicios públicos.	1. Establecer una conceptualización general de la Inteligencia Artificial.
	2. Analizar la relación entre IA y AA.PP. y las transformaciones que han acontecido en esta última.
	3. Indicar los principales riesgos que acarrea la implementación de la IA en la prestación de bienes y servicios públicos.
	4. Crear un modelo ideal de análisis de riesgos del ciudadano por la implementación de la IA.
	5. Estudiar el caso de VeriPol.

Fuente: Elaboración propia

1.4. Estrategia

Para estudiar, analizar y recopilar la información acerca de lo que se ha mencionado previamente se ha optado por una investigación cualitativa sustentada en el análisis documental[4], el análisis de un servicio público que emplea la IA en su funcionamiento y la realización de entrevistas en profundidad a expertos que han vivido en primera persona los cambios que produce la IA en este entorno.

Se ha optado por una investigación cualitativa porque permite establecer un marco de estudio y análisis más amplio que el que ofrecería una investigación cuantitativa. Se han de considerar enfoques cuantitativistas para analizar estas cuestiones, pero ha de tenerse en cuenta el cómo, dónde y quién desarrolla e implementa el sistema de IA. No solo se han de emplear enfoques cuantitativistas a la hora de abordar estos temas, sino que es indispensable tener en cuenta una panorámica más amplia, como la que permite el enfoque cualitativo (Corbetta, 2010). El enfoque cualitativo permite entender de una

[4] Bibliografía especializada – ensayos, libros, artículos académicos-, artículos de prensa, información disponible en las páginas web de sitios oficiales.

forma más amplia cómo la IA se emplea en las Administraciones Públicas y las repercusiones que su aplicación tiene para los ciudadanos, puesto que permite observar mejor cómo se interrelacionan los elementos que se quieren analizar (Corbetta, 2010).

Por ello, las técnicas cualitativas que se emplearán en la investigación son las siguientes:

- Análisis documental

El primer pilar de la investigación es el análisis documental de bibliografía especializada en IA y Administración Pública, sobre todo para establecer los marcos conceptuales en los que se asientan la investigación y sus categorías de análisis.

Se ha de establecer un estudio a nivel general del nivel de la integración de la IA en la actividad de las instituciones públicas, para así atisbar cómo la IA se integra en el proceso de implementación de un servicio público y los posibles riesgos que pueden darse para el ciudadano.

- Estudio de caso

El segundo pilar fundamental de la estrategia de investigación es el análisis de un servicio público que emplea IA en su ejecución, en concreto se estudiará el diseño del servicio público VeriPol y su implementación. Se establecerán una serie de categorías de análisis basadas en los resultados del análisis documental. Además, se ha de definir y contextualizar el servicio público.

Dicho servicio ha de ser contextualizado, tanto de forma externa como interna, en un entorno determinado, así como identificar sus componentes básicos y explicar en qué se basa el mismo. Sin olvidarnos de destacar los factores característicos del servicio para que este se lleve a cabo, lo que conllevará un modelo de gestión específico para llevar a cabo el servicio. Todo ello poniendo el acento en cómo la actividad que se genera gracias a la ejecución de dicho servicio, en virtud de la IA, aporta valor público y no relega al ciudadano a un segundo plano.

- Entrevistas en profundidad a expertos

El tercer pilar, la realización de entrevistas en profundidad a expertos, consistirá en la realización de una serie de entrevistas en profundidad semiestructuradas a personas relevantes que están viviendo en primera persona los cambios que produce la utilización de la IA. En este punto también serán claves los conceptos y categorías de análisis definidos anteriormente, ya que pueden servir para identificar verbatines en sus discursos. Se ha seleccionado una estructura de entrevista semiestructurada. Dichas entrevistas serán realizadas individualmente y, no solo se buscará una posible saturación del discurso, sino que el objetivo principal de dichas entrevistas será el mostrar el uso factible o no de estas herramientas tecnológicas en la ejecución de servicios públicos sin perder de vista el beneficiario final, el ciudadano.

CAPÍTULO II. CONCEPTUALIZACÓN DE LA INTELIGENCIA ARTIFICIAL

La IA es uno de los campos más interesantes del ámbito de la computación. Una de las causas de su desarrollo fue averiguar cómo hacer que un ordenador imitara la inteligencia humana o, incluso, ir más allá y conseguir que los ordenadores pensasen por sí mismos de forma autónoma, como comenta Barrera (2012). El estudio de la IA aúna varias disciplinas desde la ingeniería informática a la lingüística y la filosofía (Barrera, 2012, p. 87) y sus aplicaciones trascienden dichos campos a áreas de estudio como la economía y la ciencia política. Por ello, realizar una breve aproximación a la historia, concepto y rasgos generales en torno a la Inteligencia Artificial permite observar las características que hacen tan particular a este tipo de tecnología y su implementación. Se ha referido en base al Libro Blanco de la UE (2020) una noción básica de la IA, pero se ha de completar dicha definición, porque el objetivo de estos sistemas, en definitiva, es simular la capacidad de razonar que posee el ser humano basándose en los *inputs* que recibe (Arenilla, 2021, p.113).

2.1. Definiciones de Inteligencia Artificial

En el primer capítulo se ha establecido que la definición que se tomará de referencia es la que se establece en el Libro Blanco de la Unión Europea (Comisión Europea, 2020b). Pero, también se ha de hacer referencia a otras formulaciones del concepto para entender cómo funciona y se ha ido construyendo este tipo de tecnología.

En una primera instancia, la IA se puede entender como el estudio del comportamiento de las máquinas, aunque esta definición propuesta por Nilsson puede resultar un tanto circular. Por otro lado, la propuesta de McCarthy es entenderla como la ciencia e ingeniería cuyo objetivo es construir máquinas o programas de computación inteligentes, y comprender la inteligencia humana a través de dichas herramientas. Por otro lado, la definición que proporcionan tanto Minsky y Shirai y Tsuji reflejaba que uno de los objetivos es hacer que la IA realice las mismas funciones que la mente e inteligencia humanas (Barrera, 2012, p. 88). Este punto no ha de pasar desapercibido, si se aplica esta tecnología a los servicios públicos es relevante, porque el diseño, implementación o mejora de dichos servicios podría realizarlos una máquina que posea inteligencia artificial en vez de una persona. Esta cuestión cobra vital importancia si lo que se busca es una mayor eficiencia en los procesos, puesto que estos sistemas superan con creces la eficacia y eficiencia que pudiera proporcionar los sistemas computacionales que se estaban utilizando, tal y como apunta la definición de IA de Russell en 2003 (Barrera, 2012, p. 88).

Todas las definiciones que se han considerado apuntan, en definitiva, a que cuando hablamos de inteligencia artificial se han de tener en cuenta dos puntos principales: la búsqueda de la racionalidad en los procesos y que dichos procesos sean desarrollados tal y como un ser humano los haría. Aunque no todas las IAs "pensarían como un ser humano" puesto que tal y como distinguen Russell y Norvig (2010) hay sistemas capaces de pensar como los humanos, otros que actúan como humanos, sistemas con pensamiento racional y otros que actúan racionalmente (Barrera, 2012, p.88).

También se han de considerar las siete cuestiones fundacionales a las que hace referencia Nuria Oliver (2021) que se plantearon en la Conferencia de Dartmouth de 1955. Dichas cuestiones delinean los principales desafíos a los que se enfrenta la IA, y que aún hoy algunas siguen fijadas como desafíos (Oliver, 2021, pp.47-48):

1. Capacidad computacional de las máquinas para realizar funciones superiores del cerebro humano y nuestra limitación de programar los sistemas para que emulen las funciones cerebrales.
2. Capacidad de adquisición de un lenguaje igual al del ser humano.
3. Capacidad de abstracción de las redes neuronales, que tienen por objetivo emular las neuronas humanas que forman los conceptos.
4. Eficiencia de resolución de problemas apropiada.
5. Superación de capacidades propias de los seres humanos y de la propia máquina. Gracias al *machine learning*, la IA aprende de sus errores "superándose a sí misma".
6. Capacidad de pensamiento abstracto que posibilite a las máquinas generar sus propias abstracciones a partir de sensaciones y datos.
7. Posibilidad de aleatoriedad y creatividad en sus respuestas, que muestre que la IA es capaz de tener en cuenta sucesos aleatorios en sus elaboraciones y resultados diferentes bajo las mismas premisas, facilitando así un pensamiento creativo.

2.2. Recorrido histórico de la IA

2.2.1. Historia de la IA general

A la hora de realizar un recorrido histórico del nacimiento, desarrollo y evolución de la Inteligencia Artificial, se ha de tener en cuenta los avances realizados en la época antigua, medieval, moderna y contemporánea -aunque, son los planteamientos contemporáneos los que reciben el nombre de inteligencia artificial- (Oliver, 2020). Entre 450 a.C. y 1995 hay una serie de antecedentes que se han de mencionar. Sócrates, Platón y Aristóteles sientan las bases del concepto de algoritmo, en concreto Aristóteles con la formalización de los silogismos (Barrera, 2012, p.89). En Ilíada, Hefesto crea mujeres artificiales que reducían su carga de trabajo, a asemejo de un robot o autómata; Herón de Alejandría escribirá un libro llamado *Autómatas*, donde se describían tales máquinas. Además, se tiene constancia de que, en la época antigua, medieval y el renacimiento se utilizaban autómatas con finalidad religiosa o lúdica (Oliver, 2020, pp. 24-25), lo que evidencia que existían unas máquinas que, aunque de forma rudimentaria, funcionaban gracias a un determinado mecanismo.

En el siglo XIV, Ramón Lull presenta su *Ars Magna* en la que se describe el *Ars Generalis Ultima*, máquina que empleaba un sistema de inteligencia artificial. Ello servirá como base para la invención por Leibniz del sistema binario y los desarrollos posteriores de Boole y Frege en el siglo XIX. En ese mismo siglo Ada Lovace creó el primer programa informático, aunque su máquina Babbage que debía procesar dichos programas no se construyó por obstáculos técnicos y por el poco respaldo político, puesto que temían que se usara en las guerras (Barrera, 2012, p.89; Oliver, 2020, pp.24-25].

Será en el siglo XX cuando la IA comience a desarrollarse tal y como la conocemos hoy en día. En 1921 el dramaturgo checo Karel Čapek introdujo la palabra "robot", que proviene de "robota" que significa esclavo, cuando en una de sus obras exploraba el concepto de persona artificial. Más tarde en 1936 Alan Turing da a conocer la máquina de Turing y en 1950 será cuando proponga su famosa prueba de Turing. Entre tanto, en 1943 Warren McCulloch y Walter Pitts propusieron el primer modelo matemático de la neurona (Abeliuk & Gutiérrez, 2021).

Será en 1956 cuando nace como tal la "inteligencia artificial", término que se acuña en la convección de Darmouth a la que asistieron John McCarthy y Marvin Minsky, entre otras grandes figuras de la IA. Ese mismo año, Allen Newell, Herbert Simon y Cliff Shaw crearon el primer programa informático de inteligencia artificial. En los siguientes años se suceden diferentes robots, chatbots y programas informáticos con mejores funcionalidades que parecían asemejarse a la mente humana cada vez más (p.ej.: Perceptrón, Unimate, Eliza y RNNs, ALVINN). En 1997 se desarrolló Deep Blue, que derrotó al ajedrecista Gasparov y ello mostró como poco a poco se construían máquinas y algoritmos capaces de enfrentarse y resolver problemas cada vez más complejos, como Alexa, AlphaGo o BERT [Abeliuk &Gutiérrez, 2021; Oliver, 2020]. Hasta llegar a nuestros días con la GPT-4 de OpenAI.

Llama la atención que al mismo tiempo que se desarrollan estas tecnologías en el siglo XX-XXI, también hay intentos de democratización de estas, es decir, se busca un libre acceso a estas herramientas [OPSI, 2017; Abeliuk & Gutiérrez, 2021]. Por ejemplo, ImageNet, permitió acceder a una gran base de datos de forma gratuita y ChatGPT permite obtener de forma ordenada una gran cantidad de información en abierto. Este hecho no ha de pasar desapercibido, puesto que no solo se busca un progreso tecnológico, sino que también se busca que estas tecnologías estén disponibles para los ciudadanos (Abeliuk &Gutiérrez, 2021).

A pesar de que también ha habido momentos en los que la IA no fue tan popular -los dos "inviernos" de la IA: los últimos años de los setenta y los últimos años de los ochenta - se ha recuperado el interés por este tipo de tecnología, siendo actualmente uno de los temas de mayor trascendencia. La sociedad está inmersa desde hace unos años en la cuarta revolución industrial, en la que también se encuentran inmersas las organizaciones tanto privadas como públicas [Abeliuk & Gutiérrez, 2021; Oliver, 2020].

2.2.2. Las cuatro olas de la IA y el servicio público

Se pueden distinguir cuatro olas en las que se observa cómo las innovaciones en la IA han impactado en las organizaciones. Kai-Fu Lee distingue cuatro estadios en función de cómo la IA ha generado una disrupción en distintos sectores y que muestra como esta está inserta en nuestra vida diaria, las organizaciones y los servicios públicos (Criado, 2021, p.352). Las cuatro olas que Lee distinguen son: la IA de internet, la IA de negocios, la IA de la percepción y la IA autónoma (Lee, 2018).

En primer lugar, la IA de internet hace referencia a los algoritmos entendidos como sistemas de recomendación. Lee calcula que esta tecnología estaba con nosotros desde hace quince años, aunque cuando se volvió popular fue en 2012. Son Sistemas capaces de analizar y aprender sobre nuestras preferencias y ofrecernos contenido adaptado a esas preferencias. La clave de estos sistemas de IA son los datos digitales a

los que tienen acceso, que se convierten en útiles una vez han sido clasificados (Lee, 2018, p.105).

También el acceso a los datos es un pilar clave en la IA de negocios. Es bastante similar a la IA de internet, de hecho, se basa en ella. Las compañías utilizan los datos ya clasificados en su propio beneficio y buscan encontrar patrones de comportamiento ocultos en ellos para el ojo humano. De forma que pueden hacer "predicciones" sobre cómo se comportará el ser humano, ya que consideran tanto los detalles más evidentes que pueden establecer una relación causa-efecto como aquellos menos evidentes que también influyen en la acción (Lee, 2008, p.109). Tanto la IA de la primera ola como la de la segunda emplean de forma masiva algoritmos sobre datos para obtener una determinada información que consigue operar transformaciones esenciales en los sectores que las emplean (Criado, 2021, p.352).

Las dos siguientes olas, la tercer y la cuarta, se encuentran aún en una fase de experimentación, mientras que las anteriores se presentan como ya consolidadas (Criado, 2021). La tercera ola, la IA de la percepción, implica un paso más en el desarrollo de la IA. No solo es capaz de analizar, sino que también son capaces de reconocer información, procesarla e interactuar levemente con el medio que le rodea. Los algoritmos de la percepción se emplean en el internet de las cosas como por ejemplo el Echo de Amazon o el reconocimiento facial de Apple en los nuevos Iphones (Lee, 2018, p. 115).

Por último, la cuarta ola, la IA autónoma. Esta última ola representa la culminación en la integración de los elementos destacados en las anteriores olas. Máquinas capaces no solo de analizar una cantidad ingente de datos, sino también de percibir y elaborar una respuesta ante ese análisis de datos. Antes solo recibían una serie de datos, después ganaron la capacidad de analizarlos para encontrar interconexiones entre ellos y ahora podrán trabajar de manera autónoma con esos datos. Este tipo de algoritmos podrá realizar esas tareas de forma repetitiva y eficaz, pero no será capaz de detectar las imperfecciones de los objetos que analizan (Lee, 2018, p.125). Aunque, seguramente la meta sea que estas máquinas puedan llegar a realizar procesos complejos, como los haría un cerebro humano. De momento, la IA autónoma es capaz de realizar procesos complejos gracias a los algoritmos de aprendizaje profundo y tener así una determinada singularidad o *singularity* (Criado, 2021, p.353).

Estas diferentes fases de evolución de la IA se han de tener presentes si se estudia su relación con la prestación de servicios públicos y los ciudadanos. Se ha de considerar si la función que realiza la IA es sólo de análisis de datos, de percepción del entorno o si puede tomar decisiones de forma autónoma – por mínimas que sean esas decisiones-. Aunque tampoco se puede olvidar, como recalca Mark Coeckelgerh (2023) que la implementación de estas herramientas puede tener una serie de sesgos ideológicos, especialmente en su diseño.

2.3. Principales rasgos y tipos de Inteligencia Artificial

Las principales clasificaciones que destacar de la IA son las distinciones que se realizan entre la IA fuerte o débil y la IA dependiente o autónoma. Además de estas clasificaciones, se pueden distinguir una serie de características comunes a todas las áreas de estudio de la inteligencia artificial que también han de tenerse presentes.

En primer lugar, se entenderá que una IA es fuerte o débil según las tareas e iniciativa que sea capaz de abordar y realizar, respectivamente. La IA débil se enfoca en una única tarea y no tiene iniciativa propia, solo ejecuta las órdenes que recibe y poco a poco, gracias al aprendizaje automático o *machine learning* procesará y entenderá de forma natural un diálogo. Estos sistemas pueden entender las respuestas humanas y decidir qué acción tomar en función de cómo hayan sido diseñadas. Por otro lado, la IA fuerte, además de las funcionalidades de la IA débil, puede enfocarse en varias tareas, tener iniciativa propia e, incluso, superar la inteligencia de un ser humano; aunque, por el momento, no hay desarrollos de este tipo de IA [Arenilla, 2021, p.115; Criado, 2021, p.352].

En segundo lugar, podemos distinguir entre una IA dependiente, aumentada o autónoma. La IA dependiente o asistida permite realizar de forma más rápida ciertos procesos, como por ejemplo clasificación de la información o líneas de ensamblaje. La IA aumentada, da un paso más y es capaz de ayudar a las organizaciones y a las personas a hacer tareas complejas. Y, por último, la IA autónoma, que aún está en desarrollo, pero que sería el siguiente paso de la IA aumentada, puesto que permitiría tanto a las personas y a las organizaciones hacer tareas que no podrían hacerse de otro modo, además, los sistemas realizarían dichas tareas de forma independiente (Arenilla, 2021, p.115).

Una de las metas de la IA es llegar a ser igual que la inteligencia humana, hecho que aún queda lejano. Por ello se ha de tener en cuenta la clasificación de la IA según su grado de competencias: IA específica, IA general y los sistemas superinteligentes. La IA específica hace referencia a aquellas inteligencias que solo pueden realizar una tarea concreta y sólo esa tarea, la cual pueden llevar a cabo mejor que un ser humano. Sin embargo, la IA general es capaz de mostrar una inteligencia similar a la humana flexible, adaptable, etc. Y los sistemas superinteligentes, serían aquellos que superan la inteligencia humana, cuestión que plantea el filósofo Nick Bostrom [Oliver, 2020, p.58; Cortina, 2019, p.385].

Las clasificaciones comentadas no impiden encontrar una serie de características comunes entre estos ámbitos de estudio de la IA. Luger & Stubblefield (Barrera, 2012, p.88-89) proponen ocho características emergentes que pueden ser aplicadas a todas las áreas de estudio de la Inteligencia Artificial:

Tabla 2. Características emergentes de Luger & Stubblefield

1	Uso de ordenadores para realizar cualquier tipo de inferencias (razonamiento simbólico, reconocimiento de patrones o aprendizaje).
2	Énfasis en los problemas que los algoritmos no pueden resolver. Esto pone de manifiesto la importancia de la búsqueda heurística como técnica de solución de problemas.
3	Solución de problemas usando información no consistente y el uso de algún formalismo representacional que permita al programador compensar esta deficiencia.
4	Razonar sobre las características cualitativas significativas de una situación.
5	Intento de manejar la significación semántica y las formas sintácticas.
6	Elaboración de respuestas suficientes, aunque no son exactas, ni óptimas.
7	Uso de una gran cantidad de conocimiento específico a un dominio en la solución de problemas.
8	Uso de metaconocimiento para desarrollar sofisticadas formas de control sobre las estrategias de solución de problemas.

Fuente: Barrera (2012) y elaboración propia

Por tanto, uno de los principales objetivos de la IA es analizar y enfrentarse a situaciones y problemas de forma más rápida y eficaz de la que lo haría un ser humano. La IA es una herramienta que con su base de datos dada y su análisis resulta ser una gran aliada para ayudar a resolver problemáticas complicadas en un corto plazo de tiempo, proporcionando una respuesta rápida, eficaz y efectiva. La misma tarea de análisis que a un miembro de la Administración Pública puede suponer tres días de trabajo, estos sistemas pueden realizarlo en tres horas o, incluso, tres minutos. Por lo que la reducción de coste de tiempo es considerable y, en principio, la calidad de la respuesta proporcionada sería óptima [Ramió, 2019; Arenilla, 2020].

2.4. Paradigmas de la Inteligencia Artificial

Teniendo presente las diversas definiciones aportadas y las cuestiones fundacionales señaladas se puede distinguir entre dos grandes paradigmas: el paradigma de la IA simbólica y la IA conexionista. Ambos paradigmas emplean técnicas similares a la hora de plantear un desarrollo de la IA, pero tanto sus objetivos como su planteamiento son diferentes (Borrajo et. al, 1997).

En primer lugar, el paradigma simbólico, también conocido como el paradigma de ejecución. Su objetivo es crear sistemas cuyo comportamiento sea parecido o igual al comportamiento de una persona. Se pone el foco en el análisis y diseño de sistemas que sean capaces de solucionar problemas difíciles intelectualmente. Dichos sistemas podrán o no caer dentro de la definición y conceptualización de "inteligencia", pero sí han de considerarse amplificadores de la inteligencia (Borrajo et al., 1997, pp.80-92).

En segundo lugar, el paradigma conexionista, también conocido como el paradigma de emulación, puesto que su objetivo es emular al cerebro tanto en su funcionamiento como en su estructura. Se centra en el desarrollo y estudio de sistemas inteligentes que intentan alcanzar nuevos niveles de complejidad en las áreas de conocimiento. Es decir, poco a poco se busca que el sistema sea capaz de establecer relaciones más complejas entre los elementos que se encuentran en su base de datos. Se realiza un procesamiento simbólico de la información y se manipulan los símbolos de acuerdo con las reglas que se han establecido. Pero plantean que no se ha de olvidar que dichos sistemas interaccionan con aquello que les rodea, trabajan con la información que reciben del entorno y proporcionan respuestas apropiadas ante lo que se presente ante ellos (Borrajo et al., 1997, pp.80-92).

Considerando estos dos paradigmas se puede establecer una diferenciación entre varias subáreas en la inteligencia artificial. Se ha de distinguir entre el aprendizaje profundo, lo que se llamaría como tal inteligencia artificial y el aprendizaje automático. El aprendizaje profundo, relacionado con la IA conexionista, comprende las redes neuronales capaces de aprender a partir de grandes cantidades de datos. Dentro de lo que se llamaría inteligencia artificial, encontramos la IA simbólica junto a las bases de conocimiento. Este segundo ámbito engloba al mencionado anteriormente, así como las características de dichos sistemas, a lo que se les añade la capacidad de razonar, adaptarse y actuar. Por último, el aprendizaje automático engloba las dos subáreas anteriores. El aprendizaje automático se basa en algoritmos cuyo rendimiento mejora poco a poco cuando el sistema es expuesto a una mayor cantidad de datos (p.ej. árboles de decisión) (Abeliuk y Gutiérrez, 2021, p.18). Lo anterior expuesto se refleja en la siguiente imagen:

Imagen 1. Relación entre las diferentes subáreas de la IA

Figura 2. Diagrama de Venn que muestra la relación entre distintas subáreas de la inteligencia artificial.

Fuente: Abeliuk y Gutiérrez (2021, p.18)

2.5. Principios regidores de la IA

Se han establecido una serie de principios deontológicos que esta tecnología ha de seguir, puesto que a pesar de los posibles beneficios que pueda reportar también pueden presentarse una serie de riesgos a la hora de emplearse en diversas tareas, más aún si hablamos del ámbito público. Se han de tener en cuenta los principios clásicos de la robótica de Asimov (1984), así como los 23 principios Asilomar (2016) y los principios clave que establece la Unión Europea (AI HLEG, 2019) y que se concretan en los principios FATEN [AI HLEG, 2019; Oliver, 2020),

Las tres reglas de la robótica básicas de Asimov se han de tener siempre presentes a la hora emplear cualquier sistema que pueda desarrollar una cierta inteligencia, ya que se han convertido en reglas clásicas de la IA[5]. Al igual que sucede en el caso de los robots, la IA no puede hacer daño a un humano o dejar que éste sufra un daño. Además, debe de obedecer las órdenes que se le hayan dado, excepto cuando estas impliquen un perjuicio para el ser humano. Asimismo, el robot o inteligencia artificial debería de proteger su propia existencia respetando la protección al ser humano y las órdenes recibidas (Asimov, 1984). Esto último sería improbable que sucediera, a no ser que la IA fuerte se desarrollara plenamente.

Por otro lado, los Principios de Asilomar (Asilomar, 2016) a los que hacen referencia tanto Adela Cortina (2019) como Nuria Oliver (2020) recogen de forma bastante amplia lo que hay que tener presente a la hora de desarrollar sistemas con

[5] Las Tres Leyes de la Robótica: 1. Un robot no debe dañar a un ser humano o, por su inacción, dejar que un ser humano sufra daño.; 2. Un robot debe obedecer las órdenes que le son dadas por un ser humano, excepto cuando estas órdenes se oponen a la primera Ley.; 3. Un robot debe proteger su propia existencia, hasta donde esta protección no entre en conflicto con la primera o segunda Leyes. (Asimov, 1984, p. 6). Véase Anexo 1.

Inteligencia Artificial – y que sigue la línea de los Principios de Asimov-. Estos principios fueron propuestos en la Conferencia Asilomar por el *Future of Live Institute* en 2017. En dicha conferencia participaron grandes personalidades de la innovación científica y tecnológica que acordaron 23 recomendaciones a seguir (Oliver, 2021, p. 122). Estos principios consideran la IA como una herramienta valiosa y que es diseñada para el bien (Cortina, 2019, pp.381-382). El seguimiento de estos principios ayudará a empoderar a las personas a las que dichas herramientas van dirigidas. Cada vez se utilizan más estos algoritmos y cada vez estarán más presentes en nuestras vidas. Los 23 principios se establecieron para garantizar el uso beneficioso de la IA (Asilomar, 2016)[6] y se establecen en torno a la definición de los temas de investigación de la IA, su ética y valores, y las prospectivas de futuro.

Ha de tenerse presente tanto los objetivos de las investigaciones, como la financiación, la cultura que se establezca en torno a ello, la relación que se dé entre la política y la ciencia, teniendo presente la amenaza que estas investigaciones pueden suponer para la seguridad. Tanto las investigaciones que se realicen como al ejecutar dicha tecnología hay que tener en cuenta la ética y los valores de esta. Es decir, se ha de ser capaz de asegurar que la IA es fiable a lo largo de su vida útil y que todos sus procesos sean transparentes, claros y entendibles. Los diseñadores y desarrolladores de estos sistemas también han de responsabilizarse de sus creaciones, las cuales han de mantener sus valores alineados con los valores humanos bajo los que se diseñan (Asilomar, 2016).

Además, cobra vital importancia la privacidad personal de los datos de los ciudadanos que las IAs de los servicios públicos empleen, puesto que ello tampoco debe de limitar su libertad e intimidad. Todo ello sin minar el beneficio y prosperidad que la IA ha de proporcionar. A lo anterior se suma que los humanos han de seguir controlando dichos sistemas y ha de evitarse una carrera armamentística (Asilomar, 2016). Asimismo, se ha de tener presente la visión a largo plazo de esta tecnología, puesto que su evolución es constante y cada vez más ligada a las vidas de los ciudadanos. Por ello se habrán de tener claros en un futuro los límites de la IA, sobre todo de los sistemas superinteligentes. Y, aunque también los sistemas cada vez sean mejores y tengan menos errores; no ha de perderse de vista los riesgos que su empleo y ejecución pueden tener sobre los ciudadanos por impactos no deseados y la posible no consecución del bien común, para el que en un principio dicho sistema ha sido creado (Asilomar, 2016).

En una misma línea se asientan los principios que establece la UE y su concreción en los principios FATEN que menciona Nuria Oliver (Oliver, 2021). Dichos principios se establecen en el marco de la Estrategia para la Inteligencia Artificial de la Comisión Europea y que se publicaron en 2019 para establecer una guía ética básica de la IA (Oliver, 2021, pp.128-129). Los cuatro pilares que articulan los principios FATEN son: prevención del daño, equidad, respeto a la autonomía y explicabilidad (AI HLEG, 2019, pp.14-17). Estos cuatro pilares plantean que los sistemas de IA no han de causar daños o desencadenar situaciones que perjudican gravemente a los seres humanos. Su desarrollo y diseño ha de partir de una base equitativa y asegurar que ningún grupo social, en todas las fases del sistema de IA, sufre sesgos injustos. A lo que se suma que estos sistemas no han de tener la capacidad de manipular, condicionar o coaccionar la toma de decisiones humana. Son herramientas útiles para facilitar la toma de decisiones, pero en ningún caso ha de tomar la IA la decisión final. A veces es complicado explicar cómo funcionan los sistemas de IA a los usuarios finales y al público en general, de ahí la necesidad de que

[6] Véase Anexo 2.

dichos procesos sean transparentes, se revise el análisis de datos e información que realizan y que siempre se comprueben los resultados que presenten.

El punto de partida de la elaboración de estos principios éticos son los ciudadanos, ellos se encuentran en el centro. La IA se considera como una herramienta beneficiosa, siempre y cuando respete los derechos fundamentales de los ciudadanos. En caso de que su diseño, desarrollo y aplicación no los respetase, no sería considerada una IA fiable (AI HLEG, 2019). En los siguientes apartados se volverán a mencionar estos puntos clave y los principios FATEN, puesto que son las directrices para tener en cuenta si se sigue el análisis ALTAI (*Assessment List for Trustworthy AI*) que propone la Comisión para generar una IA confiable. La Comisión Europea seleccionó a un grupo de expertos (Comité de Expertos de Alto Nivel en Inteligencia Artificial o AI HLEG) en 2018 para analizar las implicaciones que la IA podía tener en nuestras vidas, poniendo el foco en las consecuencias sociales, legales y éticas de esta tecnología (Oliver, 2021, p.122).

Los principios FATEN se articulan alrededor de la intervención y supervisión humana de los sistemas de IA. Estos sistemas han de mostrar robustez y seguridad, así como asegurar la privacidad y una buena gestión de los datos que utilizan. Tanto el diseño, como la implementación y todo aquello que rodea han de ser transparentes, no discriminar, respetar el bienestar social, y en caso de ser posible, realizar rendición de cuentas. Los principios FATEN tal y como los describe Nuria Oliver (2021, p.122) son:

1. Intervención y supervisión humanas. Los sistemas implementados de IA no han de limitar la autonomía humana y han de permitir el desarrollo de las sociedades democráticas.
2. Robustez y seguridad. Los algoritmos empleados han de ser capaces de resolver errores a lo largo de la vida útil de la IA con la máxima seguridad.
3. Privacidad y gestión de datos. Los datos utilizados no han de perjudicar a los ciudadanos y estos han de poder inspeccionar y vigilar sus datos.
4. Transparencia. Los sistemas de IA han de ser trazables.
5. Diversidad, no discriminación y equidad. Se ha de garantizar la accesibilidad a estos sistemas y estos han de tener en cuenta la diversidad presente en la sociedad.
6. Bienestar social y medioambiental. Las IAs han de promover el cambio social positivo y ayudar en la sostenibilidad del medio ambiente.
7. Rendición de cuentas. Se han de establecer mecanismos que aseguren una rendición de cuentas responsable de los sistemas empleados.

CAPÍTULO III. La transformación de las Administraciones Públicas en la prestación de los servicios públicos

A lo largo de los años, las AA.PP. han cambiado la forma de estructurarse y prestar servicios públicos. Para poder ofrecer bienes y servicios públicos en adecuación con los intereses y necesidades de los ciudadanos, las instituciones públicas han de obtener la información necesaria para satisfacerlos (Criado, 2021). Esta información es obtenida a través de una serie de herramientas que han ido evolucionando con el paso del tiempo y la gestión que se realiza de esa información es de suma importancia (Filgueiras, 2021). Además, las transformaciones organizativas que han sufrido las instituciones públicas también traen consigo diferentes modelos teóricos ideales administrativos (Corvalán, 2017). Es importante considerar que las AA.PP. han de tener un esquema organizativo flexible que les permita asumir los cambios que acontecen en su entorno externo e interno (Muñoz, 2020) y es que una mejora en los métodos de recolección de información mejora la obtención de los datos y su calidad. Las AA.PP. han sufrido una gran transformación, de una recogida de datos rudimentaria, a la Administración electronificada con las TIC y, actualmente, avanzando hacia una Administración smartificada con el empleo de la IA (Ramió, 2019).

3.1. Evolución de la tecnología en las AA.PP. y los modelos de administración

La administración ha empleado diferentes herramientas a la hora de su gestión y prestación de servicios públicos a lo largo de los años. Es muy importante la revolución que han marcado las TIC, que electronificaron la AA.PP., y ahora la Inteligencia Artificial, "smartifica" la Administración (Ramió, 2019)[7]. Los modelos de Administración que conocíamos se han de adaptar también a estos nuevos tiempos, han de converger hacia un modelo de administración holístico. Sin embargo, no se ha de olvidar lo que comenta Fernando Filgueiras (2021, p.25):

> *"Los sistemas de IA en la administración pública representan tecnologías de poder, es fundamental cuestionar no solo su carácter técnico, sino también la legitimidad de la herramienta para resolver los problemas que afectan a la sociedad. Este proceso involucra entonces no solo las cuestiones técnicas, sino también políticas."*

La tecnología de la IA no solo implica una revolución en la forma de hacer las cosas, sino que también redimensiona el modo en el que nos relacionamos con las instituciones públicas y ellas con nosotros. Son herramientas que emplean una gran cantidad de información y datos para elaborar sus predicciones. Las Administraciones, en los últimos años, han incorporado una serie de herramientas que han ayudado a la ejecución de sus procesos y prestación de servicios (Corvalán, 2017). Estas mejoras, en

[7] "La smartificación de la Administración pública implica actualmente la utilización del *big data* con tres objetivos básicos: a) mejorar la calidad de los servicios a los ciudadanos; b) mejorar la inteligencia institucional para incrementar la capacidad de toma de decisiones, de control y evaluación de las políticas públicas; y c) mejorar la inteligencia institucional para lograr la mayor capacidad para ejercer el papel de dirección de las complejas redes de gobernanza públicas-públicas…. y público -privadas" (Ramió, 2019, p.14)

muchos casos, se deben a un perfeccionamiento en el sistema de procesamiento de la información que obtienen de los ciudadanos. Si hay una mejora en la obtención de información del entorno, se producirá una mejora en la adecuación de aquello que ofrecen las AA.PP. para cubrir las necesidades del ciudadano (Arenilla, 2021). Está aconteciendo una evolución sin precedentes a la hora de procesar datos, puesto que antes era solo nuestro cerebro quien los procesaba (Corvalán, 2017, p.28).

Este tipo de tecnologías reconfiguran, en definitiva, la organización del poder estatal y su relación con la ciudadanía. Hemos de asegurarnos que las nuevas tecnologías que se implementen en las instituciones no minen los derechos de los ciudadanos, pero tampoco pierdan su efectividad. Y es que al final, la aplicación de estas herramientas supone una transformación de nuestros sistemas políticos (Corvalán, 2017, p.29-30).

Se han de señalar dos transformaciones[8] de las AA.PP. para el caso a examinar: la evolución en los paradigmas de la administración y la evolución en los modelos ideales de administración. Los cambios que acontecen en los paradigmas y modelos de las instituciones públicas van de la mano, de una administración analógica, a una administración electronificada que se convertirá en una administración smartificada (Ramió, 2018). Muñoz (2020, p.17) comenta que Corvalán distingue entre cuatro paradigmas distintos de la Administración y la tecnología que emplea, tal y como se ve en la siguiente tabla:

[8] También cabría señalarse la transformación de los sistemas políticos, puesto que ello sería de interés y que nos daría una visión integral de lo que acontece en las organizaciones públicas. Sin embargo, ello excede los límites de este trabajo. Pero, no podemos olvidar está reflexión y la necesidad de establecer una línea de investigación en torno a esta cuestión. Véase Coeckelbergh (2023) y Lee (2018).

Tabla 3. Paradigmas de la Administración Pública y tecnología empleada

Paradigmas de la administración	**Época**	**Tecnología empleada para desempeñar su actividad**	**Tipo de burocracia**
1.0	XVIII- XX	Papel, imprenta y máquina de escribir.	Burocracia de papel
2.0	XX	Ordenador, procesadores de texto, fax e impresora.	Burocracia digital
3.0	XXI	Internet y aplicaciones móviles. Principios rectores: optimización, facilitación y simplificación.	Burocracia digital
4.0	XXI	Sistemas de inteligencia artificial.	Burocracia de papel, digital e inteligente

Fuente: Elaboración propia a partir de Corvalán (2017)

Actualmente, las organizaciones públicas se encuentran en la transición de una Administración 3.0 a la Administración Pública 4.0 [Corvalán, 2017; Ramió, 2019; Arenilla, 2020], en la que no solo se desarrolla una burocracia en papel o digital, sino que adquiere el rasgo de "inteligente". No sólo es posible detectar los problemas en la prestación de los servicios públicos de una forma más rápida, sencilla y eficaz, sino que además de detectar dichos problemas, se pueden prever y establecer diferentes estrategias de actuación frente a los posibles problemas de forma concreta y especializada (Arenilla, 2021). La burocracia racionalista de Weber (1947) se hace cada vez más precisa con estas nuevas técnicas.

Estos cambios también repercuten en el modelo de la organización de las instituciones. Por ejemplo, en el modelo burocrático (Hall, 1963), que propone una estructura y normas organizativas que producen un discurso tecnocrático que tiende a ser conservador (Ramió, 2018, pp. 28-29), la IA refuerza la tecnificación y tecnocratización de los procesos burocráticos haciendo posible una mayor trazabilidad, transparencia y control de los pasos seguidos en los procesos. Por otro lado, en el modelo gerencial o nueva gestión pública (Hughes, 1996), se busca la eficiencia del modelo burocrático y la transformación de las organizaciones, por lo que la IA encajaría de una forma satisfactoria en este modelo de organización, porque sería capaz de transformar las instituciones por completo (Ramió, 2018). En el caso del modelo de gobernanza (Peters y Pierre, 2005), en el que se pone énfasis en la estructura de redes y conexiones entre los distintos elementos de la Administración, la IA sería capaz de favorecer y dinamizar esas

interrelaciones. Sin embargo, hay que prestar atención a cómo se planificaría su gestión en el binomio público- privado, así como, las ideologías que puedan derivarse del uso de estas herramientas tecnológicas, que aún son difusas (Ramió, 2018)[9].

Tanto Miquel Salvador y Carles Ramió (2020), como Manuel Arenilla (2021) y Fernando Filgueiras (2021) comentan la necesidad de renovar la concepción de las organizaciones públicas hacia un modelo holístico. Comenta Ramió (2018) que el modelo burocrático, la nueva gestión pública y el modelo de gobernanza pueden convivir conjuntamente sin problemas y que, de hecho, es beneficioso que se den de forma conjunta. Si bien sería necesario un enfoque holístico de la organización, la sociedad y el sistema político, en el que se conciba al ciudadano y a estas tecnologías como agentes del cambio, ya que los tres presentan insuficiencias a la hora de asumir estas novedosas tecnologías (Ramió, 2018). No obstante, estos modelos defenderían una gestión pública de la IA y los datos e información que emplean para su funcionamiento. Aunque se pudiera pensar que una gestión privada fuera eficaz, como la que plantearía el modelo de estado regulador (Villanueva, 2020, pp.218-221), trasladar la gestión de estas tecnologías a manos privadas no sería una opción viable. Son herramientas tecnológicas que afectan a derechos fundamentales de los ciudadanos, por lo que esta tecnología debería de ser gestionada en el ámbito público, aunque el sector privado pueda realizar aportes significativos (Coeckelbergh, 2023).

Teniendo presente lo que se ha comentado, la estrategia a seguir por parte de las organizaciones públicas ha de ser una estrategia proactiva (Ramió, 2019). Se ha de esbozar y aplicar una estrategia que conciba a las instituciones como organismos abiertos y adaptativos al entorno interno y externo, así como a sus cambios. Plantear una estrategia proactiva conlleva aceptar y asumir como propios los cambios tecnológicos que acontezcan, manteniendo la innovación y la creatividad como puntos clave en el desarrollo de bienes y servicios públicos que favorezcan a los ciudadanos (Ramió, 2018). Adaptarse a los cambios y saber integrarlos en la cultura organizativa de las Administraciones Públicas de forma satisfactoria es clave para originar un considerable valor público en la sociedad y en su prestación de bienes y servicios públicos (Ramió, 2018).

Sin embargo, aunque se tengan presente los cambios que acontezcan en el entorno de las organizaciones, estas no han de perderse en la liquidez de valores, tendencias y variaciones que puedan darse en la sociedad (Bauman, 2016). La transformación que ha acontecido en los últimos años en las organizaciones públicas hace que se esté frente a una burocracia y una administración pública digital, inteligente y predictiva que es capaz de acelerar las tareas de gestión y toma de decisiones de forma significativa (Corvalán, 2017). Por ello Ramió (2018) recalca que en la nueva administración pública ha de primar la calidad institucional y desplegar una inteligencia institucional para poder dirigir públicamente los sistemas complejos de gobernanza a los que se harán frente en los próximos años.

[9] Porque, aunque aparentemente este modelo no haya una ideología dominante, no se han de perder de vista estas cuestiones (Ramió, 2018).

3.2. La gobernanza del dato y de la información

Como se ha comentado, uno de los elementos para tener en cuenta en la transformación de las Administraciones es la obtención de información de los ciudadanos para ofrecer y adaptar los bienes y servicios públicos que demanden y necesiten. Una buena gobernanza de datos es necesaria para una buena gestión de los servicios y protección de los ciudadanos frente a los peligros y retos que supone la IA en el ámbito público. El Libro Blanco sobre la IA de la UE (2020) remarca el interés público de estas cuestiones. La IA permite reducir los costes en la prestación de servicios y ofrecer una mejor adaptación de estos a las necesidades de la ciudadanía, y además, permite una mayor sostenibilidad en los bienes y servicios públicos ofrecidos y una mayor protección en temas de seguridad. Gracias a su capacidad de análisis de datos, que es capaz de obtener y producir, genera un ecosistema de confianza y excelencia pública para los ciudadanos inigualable, según la UE.

Tal y como comentan Salvador y Ramió (2020) la revolución 4.0 de la IA y la robótica ponen en el centro la información y el conocimiento como aspectos de gran importancia en la actuación pública. Las instituciones públicas necesitan información sobre los ciudadanos para poder adaptar sus actuaciones a sus necesidades. Miquel Salvador (2021) define *data governance* citando a Weber et al. (2009) y Khatri y Brown (2010) como: "la gobernanza de datos… establece derechos y responsabilidades en la toma de decisiones sobre la gestión y el uso de datos" (Salvador, 2021, p. 24). Miquel Salvador y Carles Ramió (2020) enfatizan que los datos son un activo clave e indispensable en las organizaciones, por ello han de establecerse una serie de responsabilidades en torno a su calidad y su gestión. Esta gestión tiene asociado una serie de derechos y deberes en los que se ha de velar por la calidad de los datos y su uso adecuado.

Si se plantea de esta forma hacemos referencia al propio ámbito organizativo de las instituciones (Arenilla, 2021), pero se han de recordar las dos dimensiones organizativas de la *data governance*: su empleo en los objetivos de la organización para la mejora de obtención y gestión de datos en la toma de decisiones y su empleo en la estructura de la organización puesto que facilita la designación de la autoridad formal; y la distribución de responsabilidades en los diferentes ámbitos. Respecto a estas cuestiones Arenilla (2021) señala que el gobierno francés recalcó la necesidad de construir una infraestructura sólida de gestión de datos con inversión pública permanente y una financiación adecuada, así como la creación de una infraestructura que refuerce diferentes resortes como el de la gobernanza (Etalab, 2014). Con la creación del Etalab se quería favorecer una gestión de los datos de los ciudadanos respaldada por el sector público.

Por ello es importante definir estrategias en el almacenamiento y tratamiento de datos y de información. Se puede distinguir varios modelos de almacenamiento y tratamiento de datos como los que considera la OCDE (Arenilla, 2021) y los modelos B2G (HLEG BGDS, 2020), G2B (ESSC, 2020) y G2G (Fan, Zhang y Yen, 2014) que comenta Manuel Arenilla (2021, pp.210-220). La OCDE distingue entre los modelos nórdicos, anglosajones y continentales (Almunia & Rey-Biel, 2020). En los primeros, los diferentes registros administrativos recopilan los datos y los Institutos Nacionales de Estadística son los encargados de gestionar dichos datos. En los segundos, se prima la descentralización de los datos ya que son las instituciones quienes elaboran y gestionan

su propio sistema de datos. Mientras que, en los terceros, se han creado instituciones específicas para el tratamiento de datos y proporcionan los datos necesarios a las diferentes organizaciones públicas en proyectos, investigaciones y solicitudes que los requieran [Almunia & Rey-Biel, 2020; Arenilla, 2021, pp.236-237].

Por otro lado, los modelos B2G, G2B y G2G[10] de administración de datos, se diferencian según el origen y destino de los datos que gestionan, ya sean públicos o privados (Arenilla, 2021). Los datos B2G, *bussiness to goverment*, son datos de titularidad privada que se comparten con el ámbito público [HLEG BGDS, 2020; Arenilla, 2021]; los datos G2B, *goverment to bussiness*, son datos de titularidad pública que se comparten con agentes privados [ESSC, 2020; Arenilla, 2021]; y los datos G2G, *goverment to goverment*, son datos de titularidad pública que se comparten entre organismos públicos [Fan, Zhang & Yen, 2014; Arenilla, 2021]. Este intercambio de registros ha de tenerse presente, ya que facilita o complica la gestión de la información según sea el servicio y bien público que se preste y los derechos fundamentales que se vean afectados por la gestión de dichos datos y la prestación del servicio (Coeckelbergh, 2023).

Por ello, la gobernanza del dato, de la información y del conocimiento que sintetiza Miquel Salvador (2021) han de ocupar un lugar principal a la hora de diseñar una buena gobernanza del dato y de los sistemas de IA. Salvador (2021) comenta que es imprescindible tener una estrategia adecuada que establezca qué tipos de datos son necesarios tratar y de qué fuentes han de obtenerse. A ello se suma el diseño e implementación de una buena arquitectura e infraestructura de datos para desarrollar los sistemas de IA y la internalización o externalización de algunas infraestructuras.

Se ha de tener en cuenta que en la estructura y procesos de la gobernanza de datos puede producirse la descentralización o centralización del almacenamiento, gestión y uso de datos. Estos procesos no podrían llevarse a cabo sin la gestión apropiada del talento y las competencias necesarias en la organización correspondiente para el buen tratamiento de los datos. La administración holística y relacional que se planteaba en anteriores apartados es clave para regir las interacciones entre todos los agentes que participan en el desarrollo e implementación de los sistemas de IA.

De esta forma, se puede establecer un marco de análisis adecuado para proyectos propuestos que empleen la IA, poniendo en el centro el énfasis en el ciudadano y una buena gobernanza (Salvador, 2021). Se ha de concebir la regulación sobre la IA, en definitiva, como una extensión de las regulaciones que se establecen para los datos (Comisión Europea, 2020a). Sin embargo, deberían tenerse presente las conclusiones que extraen Willems et al. (2022) acerca de la paradoja de la privacidad, puesto que a pesar de ser la privacidad de los datos personales una de las principales preocupaciones, los ciudadanos comparten sus datos públicamente de forma constante; tampoco parece que preocupe a los ciudadanos compartir sus datos en el uso de aplicaciones que empleen IA en los servicios públicos, pero a la vez es una de sus principales inquietudes [11].

A la hora de realizar una buena gobernanza de la IA, de los datos, de la información y del conocimiento se han de tener presente todos los elementos que inciden en ellas, especialmente los beneficios y riesgos a los que estas pudieran dar lugar.

[10] Véase European strategy for data European Parliament resolution of 25 March 2021 on a European strategy for data (2020/2217(INI)).
[11] Véase Willems et al. (2022).

3.3. Aplicación de la IA en los servicios públicos

Las Administraciones Públicas han ido adaptándose a los cambios del entorno con el paso de los años y, con la implementación de la IA en sus bienes y servicios públicos, uno de los principales focos es la buena gobernanza de los datos que estas recopilan, como se ha comentado. Sin embargo, se ha de especificar cómo se introduce la IA en la prestación de dichos servicios públicos puesto que pueden identificarse una serie de capacidades y ámbitos de aplicación específicos (EY, 2020).

La IA necesita de una gestión de datos eficaz que haga posible un fácil acceso a los datos que necesita para su funcionamiento y que dichos datos sean de alta calidad. Ello no puede darse sin una gestión del talento en las organizaciones públicas, las cuales han de tener presente el avance tecnológico, siendo conscientes de la importancia de la ética y cultura de las organizaciones públicas (EY, 2020)

En cuanto a las funcionalidades que puede aportar la IA al sector público se ha de destacar el impulso y dinamismo que ofrece a las organizaciones en sus procesos, ya que agiliza el proceso de toma de decisiones de políticas públicas o ayuda a analizar de forma rápida y eficaz las solicitudes presentadas para una determinada ayuda social (Arenilla, 2021). Asimismo, es un mecanismo eficaz en la prevención de detección de fraudes o de pandemias sanitarias, puesto que ayuda a crear modelos predictivos en base a los datos que se le suministra, lo que la hace capaz de detectar y anticiparse a situaciones que pueden suponer un riesgo para la ciudadanía y poder elaborar así estrategias que eviten dichas situaciones. Todo ello generará una base de datos propios del sistema de IA que permitirá automatizar procesos y agilizar tanto las propias actividades de las instituciones públicas, así como la prestación de bienes y servicios (EY, 2020).

Estas facilidades que reporta el uso de la IA en el diseño e implementación de los servicios públicos tienen por objetivos: acreditar las decisiones que se hayan de tomar; aumentar la eficiencia de procesos y de decisiones; reducir los riesgos en el ámbito en el que se aplique; permitir una mayor sostenibilidad; hacer más accesible las organizaciones públicas y sus actuaciones; afianzarse como garantía de calidad de los servicios públicos y promover la igualdad, eliminando los posibles sesgos que esta tecnología pudiera generar (EY, 2020, p.35). Por ello, sectores clave de empleo de la IA son desde los servicios generales públicos, hasta el orden público, la seguridad, la protección ambiental o la educación, entre otros (Misuraca & Van Noordt, 2020). Los sectores que cuentan con una mayor adopción de esta tecnología son el transporte público (50%), la sanidad (48%), y la administración pública (36%) (EY, 2020).

Tanto en la sanidad, como en el transporte y en la propia Administración Pública ha tenido un gran impacto. Se emplea sobre todo en la personalización de los servicios, la automatización de procesos y tareas rutinarias y predicción de situaciones características. El empleo de esta tecnología en las actividades y prestaciones de estos ámbitos puede llegar a tener un impacto potencial capaz de transformar la organización entera modificando sus procedimientos tradicionales. Pero por el momento, el impacto de la IA en el sector público solo ha supuesto una mejora en su organización, aún no ha transformado por completo las instituciones (EY, 2020).

En el caso específico de las AA.PP. se ha empleado la IA para hacer más flexibles los procesos digitales y ayudar a los ciudadanos en su experiencia de usuario, la IA

permite adaptar los servicios prestados a las necesidades particulares de cada usuario. Gracias al *block chain* y al *maching learning* es capaz de almacenar toda la información y todos los pasos que se han seguido hasta llegar al resultado final. Su capacidad de cruce de datos permite elaborar una serie de patrones que detectan situaciones como "normales", y ante cualquier anomalía o variación en la "normalidad" del proceso es capaz de detectarlo y señalar posibles errores o fraudes. De hecho, la IA ha demostrado una gran eficacia a la hora de detectar fraudes (EY, 2020).

Esto supone un gran avance para las instituciones públicas, supone una mayor estructuración y ordenación de sus datos. Dada la rápida observación y comparación de los datos que estos sistemas pueden realizar, se multiplican exponencialmente las soluciones planteadas que amplían las posibilidades de resolución de los problemas a los que las organizaciones se enfrentan. Las Administraciones pueden resolver ahora problemas críticos, que antes no podían por no tener la capacidad de análisis y adquisición de información, que con estas novedosas tecnologías ahora poseen (EY, 2020, p.27).

Sin embargo, la mayoría de las organizaciones aún se encuentra en las primeras fases de implementación de la IA y no hay aún una presencia tal como para darse una transformación completa de la prestación de los servicios públicos (EY, 2020). Dicha transformación y cambio de paradigma en las organizaciones públicas no es debida al desarrollo de una IA fuerte- que aún no se ha desarrollado-, sino que la metamorfosis de las Administraciones provendrá de la digitalización masiva y mejora de los procesos de producción que permitirán ofrecer mejores servicios públicos con un menor coste para las entidades públicas (OSPI, 2017).

.

CAPÍTULO IV. RIESGOS Y BENEFICIOS DEL USO DE LA IA

Como se ha comentado en los anteriores apartados el uso de la IA en el sector público puede ser muy beneficioso, pero también acarrea una serie de responsabilidades ya que presenta riesgos relevantes, puesto que esos beneficios y oportunidades pueden volverse riesgos y debilidades. Se ha de señalar que este campo de estudio es bastante reciente en comparación con otros, y es un terreno en el que todavía hay poca evidencia empírica (Misuraca & Van Noordt, 2020). Todavía no se puede afirmar con seguridad cuáles son las consecuencias que tendrá la aplicación de la IA en los servicios públicos, por ello se ha de establecer una metodología de análisis de las consecuencias de su uso en los servicios públicos desde una perspectiva pública y de valores (Misuraca & Van Noordt, 2020), para poder señalar así con exactitud los beneficios y riesgos que podrían reportar estas tecnologías.

4.1. Beneficios de la IA

Los beneficios que reporta la IA son múltiples y variados. El empleo de esta tecnología en la prestación de servicios públicos conlleva una serie de mejoras en su prestación y que, sin duda, suponen una gran ventaja para el ciudadano. Se ha de tener presente que la innovación que representa la IA en el sector público incide de dos formas: capacidad de sobrepasar el número de propuestas creativas de los equipos humanos y ayuda a agilizar los procesos burocráticos (Campos, 2019, pp.85-86). Lo que reporta beneficios tanto interna como externamente a la organización pública.

Los beneficios que reporta el empleo de la IA en el contexto externo a la organización pueden ser muy provechosos. La ONU (2021) y el Parlamento Europeo (2022) remarcan que el uso de la IA puede ser muy beneficioso. La ONU señala sobre las TIC, que tienen el potencial de brindar nuevas soluciones a los problemas de desarrollo y crecimiento económico (Arenilla, 2021). Además, promueven el acceso a la información y los conocimientos, lo que supone una posibilidad de acelerar el progreso humano. Estas herramientas ayudarán a las comunidades a desarrollar tecnologías cada vez más inclusivas que reflejen sus propias prioridades y necesidades. También se ha de destacar su eficiencia y reducción del tiempo de análisis de información, datos y problemas, lo que también favorecería positivamente a la comunidad. Por otra parte, y en una misma línea, se sitúa el Parlamento Europeo (2022), que destacó que la IA asentará el desarrollo político, económico y social de los países que la implementen en sus instituciones. La UE considera cinco pilares de la sociedad a los que reporta resultados positivos la aplicación de estas herramientas: los ciudadanos, las empresas, los servicios públicos, la protección y seguridad y, sobre todo la democracia (Parlamento Europeo, 2022).

Se ha de resaltar que la IA se puede implementar en servicios tanto públicos como privados que abarcan desde una mejora en la atención médica, una mejor red de trasportes y facilidad de acceso a la información, entre otros ejemplos (EY, 2020). La IA supone una reducción de costes y una generación de nuevas oportunidades para los servicios públicos que ayudará en la consecución de los objetivos de la agenda 2030 (OHCHR, 2022). Ya que, además, estas tecnologías aumentan la seguridad y protección tanto de los ciudadanos y empresas como instituciones puesto que ayuda a prevenir los delitos, porque

estos sistemas son capaces de realizar un análisis y evaluación de datos y de riesgos rápido y eficaz. También supone un refuerzo en las democracias puesto que ayuda a combatir la desinformación y ayuda a la difusión de información de calidad. Cuanta más información se pueda contrastar y de mayor calidad, se establecerá una base de datos amplia que minimizará los sesgos y prejuicios (Parlamento Europeo, 2022). A la hora de pensar en los posibles beneficios se ha de ser conscientes que no hemos de buscar su utilidad en el número de problemas que sean capaces de resolver sino *en su ayuda a* resolverlos (OSPI, 2017)

Para que haya una satisfactoria repercusión en el ámbito externo a la institución, también ha de haber un uso provechoso de estas tecnologías a nivel interno. Las funcionalidades, capacidades y objetivos que se han comentado sobre la IA en el sector público pueden suponer una serie de beneficios a las organizaciones públicas que pueden transformarlas, como se ha señalado (EY, 2020, p.35). De entre los principales beneficios a destacar son: la mejora en la optimización de procesos y la transformación de la prestación de servicios e involucrar a los grupos de interés [EY, 2020; Ramió, 2019], ya que los procesos de las administraciones se hacen más eficaces y eficientes. Eficaces, puesto que se tiene de respaldo y apoyo una tecnología muy precisa con bajo margen de error. Y eficientes, porque reducen el número de recursos empleados en aquello en lo que se empleen y son capaces de manejar tanto grandes como pequeñas cantidades de datos (Ramió, 2018, p.32).

La IA fiable, tal y como la define la UE, ofrece una serie de oportunidades gracias a sus capacidades analíticas (AI HLEG, 2019). Puede estudiar los rasgos de la población y hacer un estudio de los sesgos, problemáticas y puntos fuertes de la sociedad, para así dedicar más recursos a aquellos que lo necesiten. Además, favorece la sostenibilidad puesto que se puede conocer con mayor exactitud las necesidades energéticas y su empleo en el transporte público puede mejorar su optimización y favorecer su uso entre los ciudadanos. Por ejemplo, en el caso sanitario, permite detectar patrones de enfermedades que se den más frecuentemente en una cohorte de población y pueden establecerse tratamientos personalizados a cada paciente, hecho que antes era más complicado de hacer posible (AI HLEG, 2019). La IA supone para la Administración y la prestación de servicios públicos una mejora en su eficacia, será más "inteligente" y más participativa (Ramió, 2018). Estamos ante la gestión pública 4.0 que optimiza la gestión y la gobernanza pública gracias a la ayuda de estas nuevas tecnologías inteligentes.

Para que sea posible obtener los beneficios comentados la IA ha de ser fiable (AI HLG, 2019, p.10), es decir, que cumpla con una serie de requisitos que la hagan ser confiable y para ello ha de ser lícita, ética y robusta. Se han de establecer una serie de fundamentos que rijan qué se entiende por una IA fiable, que respeten los principios éticos explicados en anteriores apartados – autonomía humana, prevención del daño, equidad y explicabilidad-. A la hora de plantear la inserción de la IA en los servicios públicos hemos de examinar y abordar las fricciones que pueda haber entre dichos principios, y que los beneficios expuestos que puede tener esta tecnología pueden volverse contra ella como se verá en el siguiente apartado. Para garantizar la implementación de la IA fiable se han de llevar a cabo siete requisitos clave, los principios FATEN, que se han comentado en el apartado 2.5 del presente trabajo, para así evaluar a la IA a lo largo de su ciclo vital mediante métodos técnicos y no técnicos.

4.2. Riesgos de la IA

El buen empleo de la IA proporciona al ciudadano y a la actividad de la prestación de servicios públicos, una serie de considerables beneficios y oportunidades. Sin embargo, no se ha de perder de vista los riesgos que conlleva el empleo de esta novedosa tecnología. Son sistemas tecnológicos capaces de elaborar una serie de respuestas, que, en algunos casos pueden parecernos mejores incluso que las propuestas por los seres humanos. Por ello, cuando se habla de riesgos de la IA se ha de hacer referencia desde el aspecto más técnico al aspecto más social. También, se ha de señalar las diferentes "trampas" que surgen al seguir las recomendaciones de estos sistemas. Además, se han de señalar una serie de posibles riesgos de la IA que toman de referencia los principios FATEN.

En primer lugar, si atendemos a los aspectos técnicos, se ha de diferenciar entre varios niveles de riesgos: riesgo mínimo, riesgo limitado, riesgo alto y riesgo inaceptable. Los diferentes niveles de riesgos son establecidos por la UE (Parlamento Europeo, 2023) en base a la peligrosidad de sus sistemas. Dicha peligrosidad se establece según dichos sistemas empleen una inteligencia artificial más o menos débil, según el nivel de interacción que establezcan con el ser humano y según el daño potencial que suponga para la integridad de los ciudadanos o una violación de derechos humanos [12].

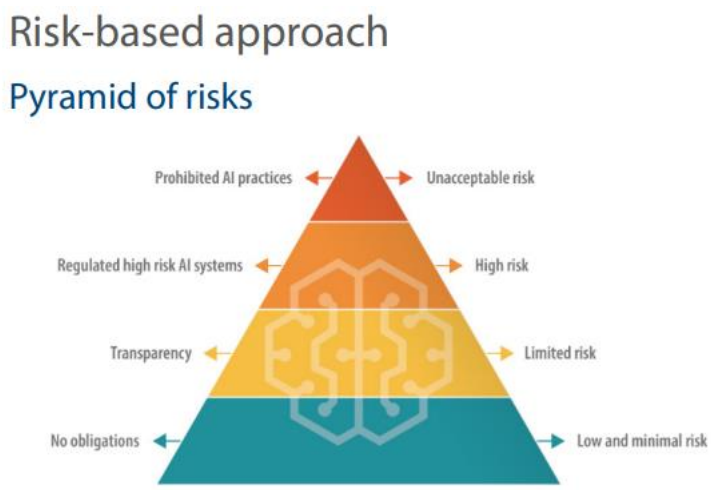
Una IA que obtenga la calificación de riesgo mínimo será aquella que no reporte ningún riesgo para el ser humano. Como no comportan consecuencias significativamente negativas no habrán de seguir un control específico, pero la UE enfatiza que sería adecuado que se adhiriera a un código ético. Si el sistema de IA empleado tiene un riesgo moderado, el sistema que se emplea interactúa con los humanos o tienen reconocimiento biométrico. Por ello, este tipo de sistemas han de someterse a una serie de controles, que, si bien no serán tan exhaustivos como los de niveles superiores, estos sistemas han de seguir una serie de supervisiones. En concreto, esas supervisiones se centrarán en la transparencia de sus procesos.

Por otro lado, las IA calificadas de alto riesgo necesitan de una regulación específica. Las IAs de alto riesgo son aquellas que se emplean como: componentes de seguridad de algún producto que se encuentre bajo la armonización de la regulación europea sanitaria y de seguridad; y los sistemas que se encuentran dentro de una serie de categorías específicas (identificación biométrica y categorización natural de personas, gestión y operaciones de infraestructuras críticas; educación; empleo; acceso y disfrute de los servicios públicos y privados y sus beneficios; refuerzo legislativo; migración, asilo y control de fronteras; administración de justicia y procesos democráticos).

Sin embargo, aún hay otro nivel de peligrosidad: el riesgo inaceptable. Los sistemas de IA a los que se les asocie este nivel de riesgo estarán prohibidos ya que sus prácticas son perjudiciales. Ejemplo de estos casos serían las IAs capaces de manipular, de realizar un reconocimiento biométrico en tiempo real, de incentivar sesgos y aquellas que las autoridades empleen deliberadamente para su beneficio personal (Parlamento Europeo, 2023, p.4-5).

[12] Véase (Parlamento Europeo, 2023).

Imagen 2. Pirámide de riesgos de la IA



Fuente: Parlamento Europeo con datos provenientes de la Comisión Europea (2023, p.4)

A pesar de la escala de riesgos comentada, la IA puede ser una gran aliada si se hace un uso responsable de esta. Sin embargo, en su uso también se ha de prestar una especial atención a los análisis que elabora. La IA puede ser un elemento clave a la hora de la toma de decisiones puesto que es capaz de analizar una cantidad ingente de datos en poco tiempo y ofrecernos una síntesis satisfactoria. Sin embargo, hemos de tener en cuentas las "trampas de las abstracciones": la trampa de los marcos conceptuales, la trampa de la probabilidad, la trampa de la formalización, la trampa del efecto dominó y la trampa del solucionismo (Misuraca & Van Noordt, 2020).

No se ha de asumir que los análisis y las comparaciones de datos, que realicen los sistemas de IA, son decisiones definitivas que reportan siempre y, en todo momento, la respuesta más adecuada. La trampa de los marcos conceptuales muestra el peligro de creer que conceptos sociales, como la equidad o la justicia, pueden ser definidos y entendidos por este tipo de sistemas.

Algo parecido sucede con la trampa de la probabilidad, porque, aunque puedan darse patrones repetitivos en ciertos contextos sociales, cada situación es única y puede no acontecer de la misma manera. El algoritmo predice lo que es más probable que suceda, pero no ha de entenderse que esas situaciones van a ser siempre así, ni tener las mismas características y la solución que en un momento sirvió para paliar una situación parecida, puede que, en la nueva situación a analizar, no sirva y haya que adaptar la solución propuesta. De ahí la importancia de que siempre un ser humano revise los análisis que llevan a cabo estos sistemas (Misuraca & Van Noordt, 2020).

Un problema similar es el que plantea la trampa de la formalización. La IA puede ayudarnos en la obtención de diferentes propuestas de solución para los problemas sociales en los que la queramos aplicar, sin embargo, un problema social no ha de resolverse mediante formalismos matemáticos. La IA ha de ayudar en el proceso de toma de decisiones, pero no ha de tomarse su decisión cómo la única posible, sobre todo si genera perjuicios a los ciudadanos (Misuraca & Van Noordt, 2020).

La trampa del efecto dominó sigue el mismo planteamiento que las anteriores. Señala el fracaso en entender cómo realmente el empleo de esta tecnología en asuntos sociales repercute en la propia organización y valores del sistema.

De este planteamiento también se deriva la última trampa, la trampa del solucionismo. A la hora de emplear estas tecnologías en la resolución de problemas, dado su alto nivel de eficiencia y eficacia, se puede caer en la trampa de pensar que la mejor solución siempre habrá de pasar por el empleo de esta tecnología, cuando a lo mejor no es necesario, ni recomendable (Misuraca & Van Noordt, 2020).

Al considerar estas "trampas" se plasma una paradoja en el empleo de la IA. Puede darse la situación en la que se siguen los consejos de las IA -resultado de sus disquisiciones-, es decir, se permite que "gobierne" un algoritmo, mientras que este a su vez es gobernado por otro algoritmo, que lo evalúa y lo controla. Se crea así un bucle infinito de algoritmos que "gobiernan" a otros algoritmos, sin los ciudadanos realmente gobernar, puesto que se ha "cedido" la responsabilidad de la toma de decisiones a estas tecnologías (Misuraca & Van Noordt, 2020).

Estos algoritmos se distinguen de otros tipos de tecnologías por sus habilidades de percepción, razonamiento y acción, capaces de crear externalidades que las anteriores tecnologías empleadas en la toma de decisiones y prestación de servicios públicos no eran capaces de generar (Misuraca & Van Noordt, 2020). A lo anterior se añade que se podría entrar en la "trampa de la perfección burocrática y perfecta racionalización" que el empleo de esta tecnología tendría en el modelo burocrático weberiano (Weber,1947).

Todo ello incide en los desafíos que se enfrentan a la hora de emplear la IA en la prestación de servicios públicos. Se han de considerar los problemas de la utilización excesiva o infrautilización de las funcionalidades de estas tecnologías, ya que sobre todo un uso abusivo puede acarrear graves problemas. Además, intentar resolver "de forma sencilla y rápida" problemas sociales complejos puede ser una decisión equivocada, a lo que se suma la distorsión económica que su uso puede generar tanto en el ámbito privado como en el público y el cambio que supondrá en el mercado laboral[13] [Parlamento Europeo, 2022; Ramió, 2019].

A la hora de emplear este tipo de sistemas ha de especificarse sobre quién reside la responsabilidad en caso de haber cualquier problema o si se producen externalidades negativas: el propietario, el fabricante o el programador. La mayoría de los sistemas de IA gestionan una gran cantidad de datos, en algunos casos, datos personales, por lo que se han de considerar los posibles problemas que se pueden generar respecto a los derechos fundamentales de los ciudadanos y la democracia, y los peligros que presenta el *mathwashing* (Allo, 2018) al considerar una tecnología como objetiva y precisa, cuando no lo es.

El impacto de la IA en nuestros servicios públicos presenta también desafíos en la seguridad y protección, especialmente, aquellos sistemas que interactúan con los ciudadanos y predicen su comportamiento. Asimismo, supone un gran reto para la transparencia tanto en su diseño, uso e implementación. El seguimiento de las acciones realizadas con estos sistemas es más fácil de realizar gracias a la tecnología de bloques

[13] Se habría de hacer un análisis más profundo de las consecuencias laborales que reportaría la introducción de IA. Aunque no se mencione este tema en detalle, ha de ser tenido en cuenta y realizar un estudio más profundo.

de los algoritmos, si bien es cierto que las decisiones de emplear una IA u otra dependerán de la agencia y supervisión humanas.

4.3. Los 7 puntos clave

Los desafíos de la implementación de la IA en los servicios y bienes públicos comentados pueden verse como oportunidades para mejorar el uso de esta tecnología. Como han de tenerse presente los riesgos a los que a analizar, se ha de establecer una guía de estudio que asegure que la IA que se emplea en los servicios públicos sea una IA fiable, antes y durante su período de servicio. Para determinar si una IA es fiable y no portadora de riesgos se ha de tener presente la autoevaluación que propone el Alto grupo de expertos en IA de la UE (AI HLEG). El grupo de expertos establece siete puntos de autoevaluación, se inspiran en las directrices éticas básicas comentadas y los principios FATEN, porque tienen como punto central la protección del ciudadano. Son puntos claves para determinar los riesgos que puede desarrollar un sistema de IA a la hora de emplearse en la prestación de servicios públicos y asegurar así que el algoritmo se ejecuta tal y como ha sido entrenado (AI HLEG, 2020):

4.3.1. Agencia y supervisión humana

A pesar de que la IA puede actuar hasta cierto punto de forma independiente, cuando se emplean estas tecnologías en los procesos de toma de decisiones siempre ha de ser supervisada por un humano. Su uso supondrá un beneficio siempre que agilice dichos procesos y genere externalidades positivas. Pero, en ningún caso ha de influenciar o manipular la autonomía[14] humana en la toma de decisiones para asegurar una IA fiable y que ayude al desarrollo de una sociedad democrática y la defensa de los derechos fundamentales.

El riesgo que corre la autonomía humana por el empleo de la IA es la influencia que esta pueda tener en las decisiones humanas. Se espera que esta tecnología se comporte y razone como un ser humano (Coeckelbergh, 2023) cuando no lo es, no tiene capacidad de decisión, sino que ayuda a decidir. En este punto el AI HLEG recalca la necesidad de que el ser humano supervise todas las fases desde la creación hasta la implementación de estas tecnologías. Son los seres humanos quienes tienen la capacidad de decisión y la habilidad de decidir cuándo usar dicha tecnología.

4.3.2. Robustez y seguridad técnica

Es un requerimiento esencial para poder tener una IA confiable. Ha de ser una IA robusta, lo que requiere tener previsto una serie de riesgos y posibles cambios potenciales que se puedan realizar en el sistema, teniendo como objetivo no solo la seguridad del propio sistema, sino la seguridad de los ciudadanos, manteniendo en todo momento la

[14] Véase Coeckelbergh (2023) en relación con la teoría del *nudge*.

transparencia en las acciones. Se han de proteger los puntos críticos que son vitales para el funcionamiento correcto y beneficioso de la IA.

Se debe tener presente que han de establecerse una serie de métricas para identificar los diferentes niveles de riesgo que puede desarrollar el sistema de IA. Es indispensable la identificación de las posibles amenazas, así como saber cuan necesario es ese sistema en la ayuda de toma de decisiones. Para que la IA sea fiable y confiable ha de tener un comportamiento estable. Por ello es indispensable que se elabore un mecanismo de evaluación de la IA empleada que alerte cuando la IA no cumple con sus correctos niveles de robustez técnica y seguridad.

Para garantizar esa seguridad es necesario que el sistema de IA también sea preciso y cumpla con los niveles de exactitud previstos, de los que también han de ser conocedores los usuarios finales. Por lo que la monitorización de este tipo de aplicaciones es clave para saber si mantiene los niveles adecuados de precisión, si sus datos están debidamente actualizados y si realiza operaciones válidas en función de cómo ha sido entrenada.

Para evitar los efectos negativos que pueden provocar la pérdida de robustez y seguridad, así como de precisión de los sistemas de IA, se han de tener previsto planes alternativos para asegurar su fiabilidad.

4.3.3. Privacidad y gobernanza de datos

Este es uno de los puntos clave a la hora de hablar de los posibles riesgos de la IA, como se ha comentado en anteriores apartados. Si no se realiza una adecuada gobernanza del dato, el riesgo del ciudadano a la hora de emplear este tipo de tecnologías aumenta. El derecho a la privacidad de los ciudadanos es un derecho fundamental que ha de respetarse, así como la integridad física y psicológica. Por ello los sistemas de IA han de respetar la ley de protección de datos correspondiente que le permita procesar los datos con capacidad suficiente sin contravenir la citada ley.

Por estas cuestiones es importante que se lleve a cabo una gobernanza del dato, en la que se tenga en cuenta lo necesario para establecer un entorno seguro y de protección al ciudadano. Habrá de tenerse en cuenta si dichas IAs han sido entrenadas con datos personales o datos generales para aplicar el tratamiento de datos correspondiente. Además, ha de reconocerse el derecho del ciudadano a que sus datos no sean usados sin su consentimiento en las bases de datos de las IA, ya sea tanto en su entrenamiento como en su ejecución; y el derecho al olvido de sus datos y revocación del permiso de autorización de utilización de estos. Si el sistema de IA no tiene en cuenta estos aspectos, así como las normativas de protección de datos nacionales e internacionales[15], el sistema no será un sistema de IA fiable y presentará un riesgo serio para los ciudadanos.

[15] En el caso de España se aplican el Reglamento General de Protección de Datos (GDPR) aprobado en 2016 por la UE que entró en vigor en 2018 y la Ley Orgánica 3/2018 de Protección de Datos Personales y garantía de los derechos digitales (LOPDGDD).

4.3.4. Transparencia

Este es otro de los puntos clave a la hora de analizar los riesgos que la IA presenta al ser empleada en la prestación de servicios públicos. El sistema de IA debe hacer accesible y posible su trazabilidad, su capacidad explicativa y su comunicación.

Respecto de la trazabilidad, se ha de señalar que las diferentes fases, desde el diseño hasta la ejecución, han de poder conocerse, ya que, por un lado, el poder explicar los procesos técnicos que ha seguido paso por paso el algoritmo, así como su monitorización, favorece la confianza en la IA. Se debe evitar la existencia de "cajas negras" que generen contextos opacos que no permitan acceder a la información necesaria para entender cómo funciona el sistema de IA empleado en toma de decisiones o prestación de servicios públicos.

El poder acceder y explicar los procesos técnicos es crucial para entender por parte de los usuarios finales y el resto de los ciudadanos cómo funciona el sistema de IA empleado. Saber cómo funciona el mecanismo que ayuda en la toma de decisiones y por qué se ha llegado a la conclusión que se ha llegado, genera una mayor confianza en la sociedad. Lo que exige que sea posible una comunicación comprensible por parte de los usuarios y que puedan entender de forma clara la articulación de los sistemas, así como sus limitaciones, que también deben ser comunicadas.

4.3.5. Diversidad, no-discriminación y justicia

Otro de los aspectos importantes es que, tanto en el diseño como en su ejecución y en la toma de decisiones y prestación de servicios, el sistema de IA debe de ser accesible para los usuarios a los que va destinada y no ha de generar situaciones discriminatorias o que favorezcan más a unos usuarios que a otros. Se ha de evitar la potencial marginación de grupos sociales por sus características específicas o la no posibilidad de acceso a estas tecnologías.

Las "trampas" que se han comentado en este mismo apartado, juegan un vital papel en el desarrollo de los algoritmos, tanto en su proceso de elaboración como en la actividad a la que están destinados, que se concreta en ayudar en la prestación de servicios públicos. Se definen de forma abstracta conceptos complejos que los algoritmos han de aplicar a situaciones sociales complejas, de forma que es posible que creen abstracciones que generen situaciones desiguales. Por ello, se han de tener en cuenta la opinión de los *stakeholders*, puesto que ellos pueden dar una retroalimentación valiosa sobre los posibles sesgos que tuvieran los algoritmos o que estos pudieran desarrollar; así como la monitorización constante por parte de seres humanos que detecten este tipo de problemas.

4.3.6. Bienestar social y medioambiental

Los sistemas de IA han de considerar el impacto en el bienestar social y medioambiental que generan sus externalidades. La IA puede beneficiar a la democracia,

a los ciudadanos y al medio ambiente, siempre y cuando se haga un uso responsable de la tecnología.

Se han de analizar los aspectos positivos y negativos del impacto de los algoritmos en el medioambiente. Su aplicación y desarrollo ha de respetar el medioambiente, por lo que estos sistemas deberían de incorporar una medición de su impacto ambiental, para así detectar cuando suponen un riesgo. Es más, en algunos casos se han de utilizar ordenadores especiales para ejecutar los algoritmos, lo que supone un impacto medioambiental mayor y más agresivo.

Además, mal empleada, contribuye al deterioro de las habilidades sociales de los ciudadanos y perjudica a su bienestar físico y mental. Asimismo, supone un impacto en el trabajo y las habilidades a desempeñar por los ciudadanos que trabajen tanto para el sector público como para el sector privado, ya que han de adaptarse a las nuevas dinámicas que surgen de la aplicación de la IA (Ramió, 2018).

Si los servicios públicos que emplean esta tecnología no tienen en cuenta el punto de vista social a la hora de confeccionarse pueden generar sesgos indeseados y desarrollar el problema del *mathwashing* (Allo, 2018). Se ha de tener en cuenta los efectos que tendrá en las instituciones públicas y en la democracia, especialmente en aquellas relacionadas con el curso democrático, como por ejemplo los procesos electorales. Por ello, los sistemas de IA que se empleen han de estar alineados con los ODS (OHCHR, 2022) y garantizar el futuro democrático a las siguientes generaciones.

4.3.7. Accountability o rendición de cuentas

Este último punto es necesario para asegurar que los otros puntos señalados se monitoricen. Para que una IA sea calificada como fiable a la hora de prestar servicios públicos ha de poder realizarse una rendición de cuentas y una gestión del riesgo adecuada, lo que también supone un ejercicio de transparencia. Ha de ser posible realizar la auditoría de estos sistemas tanto de forma interna como por parte de terceros y de forma independiente. Ello asegura la neutralidad en el análisis sobre los aspectos anteriores y asegura la fiabilidad y la confianza que habrán de tener los sistemas de IA que se emplean en la toma de decisiones y en la prestación de servicios públicos.

CAPÍTULO V. DISEÑO METODOLÓGICO

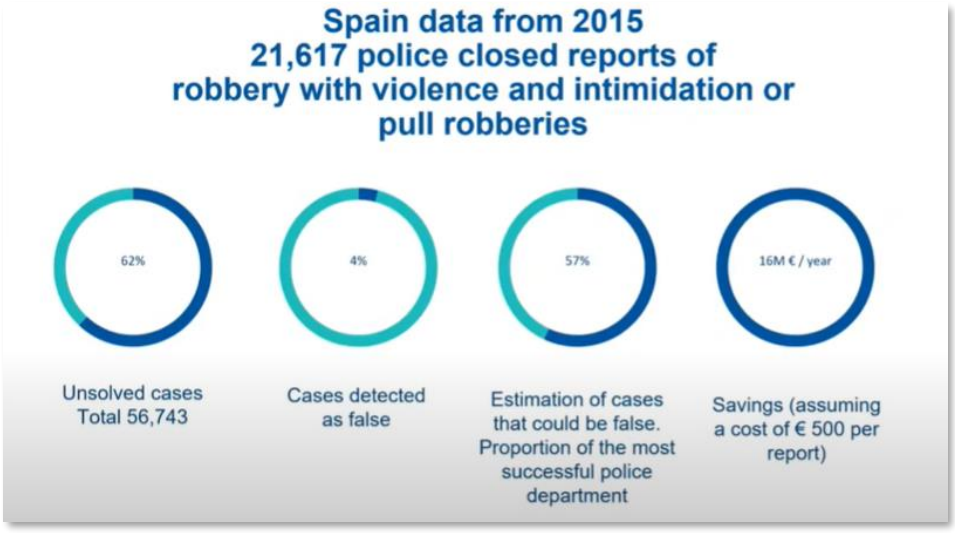
Teniendo presente lo anterior, se pueden establecer una serie de categorías de análisis que permitirán estudiar los posibles riesgos a los que se enfrentan los ciudadanos, si se decide emplear la IA en la prestación de un servicio público. El haber enmarcado el concepto de Inteligencia Artificial y algunos de los aspectos más relevantes a tenerse en cuenta, si hablamos de la relación entre IA y Administración Pública, permite establecer una guía adecuada a seguir para el estudio de un caso en concreto, que permitirá un análisis más preciso de los posibles riesgos a los que se enfrentan los ciudadanos.

5.1. Caso de estudio elegido

El caso de estudio elegido es VeriPol, la herramienta informática que emplea la Policía Nacional en las comisarías españolas para detectar denuncias falsas de robos y hurtos desde 2018. En concreto, VeriPol "estima la probabilidad de que una denuncia por robo con violencia e intimidación o tirón sea falsa" (Dirección General de la Policía , 2018). Este sistema de IA analiza el texto de la denuncia y de forma automática, sin intermediación de ningún usuario descubre si la denuncia es falsa o no. La efectividad de esta aplicación alcanza una precisión superior al 90% en la mayoría de los casos, por encima de los niveles alcanzados por un policía, un 75% (Dirección General de la Policía , 2018). Lo novedoso de este sistema de IA es que se ha trabajado y entrenado sobre documentos reales y no sobre textos ficticios creados expresamente para ese testeo (BigData_uc3m, 2022). Es automática y poco invasiva, ya que la herramienta es capaz de determinar solo mediante el análisis del lenguaje empleado en la denuncia interpuesta si esta es falsa o no (Quijano et al., 2018).

A quienes va dirigido VeriPol o quién, en definitiva, es el usuario final de esta aplicación que emplea IA, son los policías nacionales encargados de utilizar esta herramienta. VeriPol se diseña como una aplicación que pretende reducir la carga de trabajo policial y hacer más eficientes los recursos empleados en sus investigaciones. Con esta aplicación se pretende eliminar el "ruido que contamina" los registros policiales. Alrededor del 62% de los casos quedan sin resolver y la estimación de los casos que podrían ser falsos asciende al 57%. Una de las desarrolladoras de la aplicación, Lara Quijano (BigData_uc3m, 2022) recalca que gracias a la detección de denuncias falsas de VeriPol, el número de casos no resueltos descendió, y es que algunos de los casos no resueltos resultaron ser denuncias falsas. Si este sistema se aplicara en todas las comisarías podría producirse un ahorro de 16 millones de euros al año en recursos policiales que podrían destinarse a casos de mayor gravedad.

Imagen 3. Porcentaje de reportes policiales cerrados por violencia e intimidación en 2015



Fuente: BigData_uc3m (2022)

Esta herramienta es capaz de distinguir una serie de patrones en el lenguaje empleado en las denuncias falsas por los denunciantes, que en algunos casos son imperceptibles para los humanos y es precisamente eso lo que hace que sea aún más eficaz que un policía convencional. Sin embargo, esta herramienta no pretende sustituir el criterio y la toma de decisiones de los policías, sino ayudar en el proceso de toma de decisiones y conformación de un determinado criterio [Quijano et al., 2018; BigData_uc3m, 2022].

Se pueden identificar seis objetivos principales que persigue VeriPol, divididos en tres bloques, en base a lo que comentan la Policía Nacional (Dirección General de la Policía , 2018) y los desarrolladores [Quijano et al., 2018; BigData_uc3m, 2022]:

Objetivos respecto al ámbito externo de la organización:

1. Ejercer de disuasión a los ciudadanos para que estos no presenten denuncias falsas.
2. Evitar la perpetración de delitos.

Objetivos respecto al ámbito interno de la organización:

3. Evitar el empleo de recursos policiales en casos falsos.
4. Optimizar los esfuerzos de la Policía en otras tareas que requieren de una mayor dedicación.

Objetivos respecto a la aplicación del sistema de IA de VeriPol:

5. Desarrollar estrategias efectivas de prevención del delito y aumentar la efectividad de las investigaciones.
6. Desarrollar un método de predicción de veracidad a aplicar en declaraciones de víctimas de delitos más graves.

Se ha elegido el caso de estudio VeriPol por varias razones que lo hacen un servicio público peculiar y que permiten analizar los posibles riesgos que presenta la IA para los ciudadanos en su empleo en los servicios públicos.

En primer lugar, el ámbito de actuación de VeriPol es España por lo que hay una mayor facilidad de acceso a los recursos de análisis, tales como los datos e información necesaria para analizar las categorías de análisis establecidas.

En segundo lugar, es la primera aplicación informática del mundo que detecta denuncias falsas de hurto y robo, lo que supone un reto tanto a nivel policial como académico (Dirección General de la Policía , 2018). Es una herramienta especial porque el análisis que realiza del lenguaje de las denuncias es un proceso automático, sin intervención humana. En un futuro se espera que estos métodos de análisis sean aplicados a otro tipo de delitos más graves y no solo a los delitos de hurto y robo (BigData_uc3m, 2022). VeriPol es considerado como el punto de inicio de lo que podría llamarse "Policía predictiva" (Ministerio del Interior, 2018), que analizando los datos del caso abierto en ese momento sea capaz de estimar los potenciales colaboradores del delito o las posibles líneas de investigación a seguir. Esta herramienta presenta una potencial aplicación en otras fases de la investigación policial, que no solo permitirá el aumento de la precisión y acierto en la resolución de casos, sino una efectividad y eficiencia de la aplicación de recursos necesarios a los casos que realmente lo necesiten (BigData_uc3m, 2022).

En tercer lugar, ha sido premiada por su desarrollo y efectividad, así como las contribuciones que puede suponer a la labor diaria de la policía. VeriPol fue galardonada en el año 2017 con el Premio de Investigación de la Fundación Policía Española, que premia la investigación y estudio en temas del ámbito policial, y aquellos estudios que detecten nuevos riesgos o propongan nuevas estrategias para preverlos. José Antonio Nieto que en ese momento era Secretario de Estado de Seguridad, remarcó la necesidad de innovar en las organizaciones para alcanzar con éxito sus objetivos (Ministerio del Interior, 2017).

En cuarto lugar, ha sido estudiada por la UE como uno de los casos de referencia de aplicación de la IA en la prestación de servicios públicos. En el estudio que se realizó en 2020 por parte de la UE (Misuraca & Van Noordt, 2020) sobre los servicios que en Europa se prestaban mediante inteligencia artificial, de las 12 aplicaciones que existen en España, eligieron VeriPol.

5.2. Categorías de análisis

Para examinar los sistemas de IA un método a seguir es el procedimiento de autoanálisis propuesto por el AI HLEG (2019)[16]. Este procedimiento de autoanálisis, ALTAI, se fundamenta en la protección de los derechos humanos y ayuda a las organizaciones a entender los requisitos que ha de cumplir una IA para que esta sea confiable y fiable. ALTAI ayuda a minimizar los riesgos que el sistema de IA haya generado y a maximizar el beneficio de uso de la IA. Esta propuesta del grupo de expertos ayuda a las organizaciones a identificar los riesgos que la IA pueda generar y así proponer y llevar a cabo las medidas activas necesarias para minimizar dichos riesgos.

Estudiar la IA bajo el enfoque de una IA confiable es clave para establecer una competitividad responsable. Así, todos aquellos servicios públicos que utilicen la IA o los ciudadanos que se vean afectados por el uso de los sistemas de IA, pueden confiar que

[16]Se seguirá esta metodología de análisis, sin embargo, conviene señalar que los miembros del AI HLEG solo *recomiendan* usar esta metodología, por lo que no son responsables de los daños que pueda acarrear su aplicación (AI HLEG, 2019).

el diseño, desarrollo y uso de estas aplicaciones son éticas, legales y robustas, en línea con las cuestiones que se han ido comentando en los otros apartados. Las categorías de análisis, por tanto, que sigue la presente investigación se basan en lo establecido en el procedimiento de autoanálisis ALTAI (AI HLEG, 2019), de forma que las categorías de análisis a seguir son:

Tabla 4. Categorías de análisis. definición y preguntas clave.

Categoría	Definición. Preguntas clave.
0.Impacto en los derechos humanos	Para analizar si la IA cumple con los requisitos necesarios de respeto y protección de los derechos humanos se ha de tener en cuenta la evaluación del impacto en los derechos humanos (FRIA). Preguntas guía basadas en la carta y Convención europea de Derechos humanos (ECHR) y sus protocolos, así como la Carta Social Europea: • ¿Los sistemas de IA discriminan potencialmente de forma negativa a personas en base a su sexo, color, raza, etnia, origen, edad, orientación sexual, etc.? • ¿Los sistemas de IA respetan los derechos de los niños? ¿Se tiene en cuenta los intereses de los niños en su protección? • ¿Los sistemas de IA protegen los datos personales de los individuos en línea con el Reglamento General de Protección de Datos? • ¿Los sistemas de IA respetan la libertad de expresión y de información y la libertad de asociación?
1.Intervención y supervisión humana	Los sistemas de IA no han de limitar la toma de decisiones humana y siempre han de servir de apoyo, no han de tomarse como la única visión para la toma de decisiones. Las preguntas guía que se establecen son: • ¿El sistema de IA está diseñado para interactuar guiar o tomar decisiones por seres humanos que afectan a otros humanos? • ¿Sus resultados pueden generar confusiones? ¿En qué medida afecta a la toma de decisiones autónoma humana? ¿Interactúa con los usuarios finales? • ¿Genera adicción o apego y hay riesgo de manipulación? ¿Hay algún procedimiento de

	seguridad que pare el funcionamiento del sistema en caso de error?
2.Robustez y seguridad	La IA ha de ser resiliente a los ataques a los que puede verse sometida, ha de poder garantizar la seguridad de su sistema. Ha de tener resiliencia a los ataques y ser capaz de proteger aquello que le rodea y se ve afectado por ella. Además, se ha de garantizar su precisión y ha de asegurarse su fiabilidad.
	Las preguntas guía que se establecen son: • ¿Se han detectado las posibles amenazas y puntos críticos? En caso de tenerlos, ¿cuáles? ¿Puede tener un uso malintencionado? ¿Tiene resistencia a los ciberataques y se han previsto las potenciales formas de ataque al sistema? ¿Se han implementado medidas para asegurar la robustez y seguridad del sistema? ¿Se ha informado a los usuarios finales de la seguridad y actualizaciones del sistema? ¿Se han identificado los diferentes niveles de riesgo para cada uso específico? ¿Si se ha analizado la dependencia del sistema de inteligencia artificial como una parte crítica en la toma de decisiones? • ¿Su comportamiento es estable? • ¿Hay mecanismos de evaluación del sistema cuando cambia la robustez técnica y seguridad del sistema? • ¿Si tiene bajos niveles de precisión da lugar a consecuencias negativas? • ¿Se han puesto medidas para asegurar que los datos que se utilizan para desarrollar la inteligencia artificial están actualizados son completos y representativos del entorno y del sistema en el que se va a desarrollar y ha tenido el entrenamiento adecuado? • ¿Hay una monitorización en todas las fases del sistema? • ¿Se ha informado a los usuarios finales de los posibles riesgos y errores del sistema? • ¿Se evalúa y se asegura que los diferentes aspectos de la IA son fiables y reproducibles? ¿Hay un plan alternativo en el caso de que esta falle o presente cualquier error?
3.Privacidad y gestión de datos	Los sistemas de IA han de garantizar la privacidad de los datos que manejan. La gobernanza de datos ha de seguir las legislaciones vigentes sobre protección de datos y establecer

	los mecanismos necesarios para evitar filtraciones de datos personales de los ciudadanos.
	Las preguntas guía que se establecen son: • ¿Se ha tenido en cuenta el derecho de la integridad física, moral y mental? ¿Se sigue el Reglamento (UE) del Parlamento Europeo y del Consejo, de 27 de abril de 2016 sobre protección de datos personales? ¿Se puede garantizar la privacidad de los datos recogidos, empleados y producidos por el sistema de IA? ¿Se puede denegar el consentimiento de la cesión de datos personales? ¿Se tiene derecho al olvido? • ¿Se entrenó utilizando datos personales? • ¿Se siguen los protocolos de gestión y gobernanza de datos adoptados en la normativa y en las instituciones?
4.Transparencia	Ha de ser posible trazar todo el diseño, desarrollo y aplicación de los algoritmos. El proceso ha de ser entendible y comunicable.
	Las preguntas guía que se establecen son: • ¿Puede establecerse una trazabilidad clara de todos los pasos que realiza el sistema en su funcionamiento? • ¿Se tiene en cuenta la retroalimentación de los usuarios finales? • ¿Se ha informado adecuadamente a los usuarios finales sobre el funcionamiento del sistema? ¿Los usuarios finales entienden el funcionamiento del sistema y sus limitaciones?
5.Diversidad, no discriminación y equidad	Los sistemas empleados de IA han de ser accesibles a todos los ciudadanos, no han de ser sesgados ni producir sesgos. Además, se recomienda la participación de los *stakeholders* en su valoración
	Las preguntas guía que se establecen son:

	• ¿Hay una estrategia para evitar los sesgos que puedan darse en el diseño, aplicación e implementación del sistema de IA? • ¿Se ha considerado la diversidad de los posibles usuarios finales? ¿Se ha establecido un plan en caso de que haya altos niveles de discriminación y sesgo? • ¿Se han establecido mecanismos de medición para ello? • ¿Las definiciones de los conceptos que emplea la IA han sido correctamente definidos? • ¿Se han valorado los riesgos que puede suponer que no todos los usuarios puedan acceder a este sistema? • ¿En su elaboración han participado las partes interesadas?
6.Bienestar social y medioambiental	El uso de la IA ha de garantizar el bienestar medioambiental conforme a los acuerdos establecidos. Además, ha de analizarse el impacto que generará su empleo en el trabajo al que se aplica. Y, sobre todo, ser conocedores del impacto que en la sociedad y en la democracia va a generar esa aplicación de la IA.
	Las preguntas guía que se establecen son: • ¿Hay aspectos positivos y negativos del impacto del sistema a nivel medioambiental? ¿Se han establecido mecanismos para la evaluación de su impacto ambiental? • ¿Tiene impacto en la forma en la que se realiza el trabajo humano? ¿Se ha consultado a instancias superiores la aplicación de este sistema? ¿Se ha facilitado a los trabajadores la comprensión del sistema mediante algún tipo de formación? • ¿Puede tener un impacto determinante en la democracia? ¿Este impacto puede ser negativo? ¿Se ha tenido en cuenta la elaboración de una estrategia para minimizar las externalidades negativas a la democracia y a la sociedad?
7. Rendición de cuentas	Ha de ser posible la auditoría por parte de terceros de los sistemas implementados de IA y ha de haber una gestión del riesgo de las herramientas empleadas que garantice la responsabilidad y rendición de cuentas.
	Las preguntas guía que se establecen son:

	• ¿Es posible realizar al sistema auditorías tanto internas como externas? ¿Las auditorías pueden realizarse por parte de terceros? • ¿Se ha establecido un protocolo propio de evaluación del sistema? • ¿Están protegidos los derechos fundamentales? • ¿Se puede realizar una revisión ética del desarrollo, funcionamiento, implementación del sistema y sus consecuencias? • ¿Se puede realizar una monitorización continua de adherencia al resto de los requerimientos antes planteados?

Fuente: Elaboración propia a partir del procedimiento de autoanálisis ALTAI (AI HLEG, 2019).

CAPITULO VI. ANÁLISIS DE RESULTADOS

Si se aplican las categorías de análisis planteadas al caso de estudio elegido, VeriPol obtiene los resultados que se presentan a continuación. VeriPol es un sistema de IA novedoso que poco a poco se irá consolidando y será más estudiado, dadas sus peculiares características. Se puede encontrar información sobre el sistema de IA que permite responder a las preguntas planteadas. Sin embargo, a la hora de realizar la investigación, puede que se obtengan más preguntas que respuestas.

6.1. VeriPol

6.1.1. Fuentes de información sobre el servicio público

Antes de describir el servicio analizado y aplicar las categorías de análisis al caso estudiado, conviene hacer referencia a las fuentes principales que han servido de ayuda para realizar la investigación. Esta cuestión es importante mencionarla puesto que en varias de las categorías de análisis se hace referencia a la accesibilidad de la información acerca de estas aplicaciones. Las principales fuentes de información son: las instituciones públicas, el ámbito académico y los medios de comunicación.

En el ámbito institucional, encontramos información sobre la aplicación VeriPol en la sección de noticias de Policía Nacional (Dirección General de la Policía , 2018), sección de noticias del Ministerio de Interior (Ministerio del Interior, 2018), una nota de prensa de la Moncloa (La Moncloa, 2018) y el estudio realizado por la UE (Misuraca & Van Noordt, 2020). Si acudimos a estas fuentes encontramos rápidamente una saturación del discurso puesto que todas las fuentes repiten la misma información con datos y palabras parecidas o iguales con lo que la Policía Nacional escribe en su página web.

La información que se encuentra en el ámbito institucional recoge de forma resumida lo que los creadores de VeriPol presentan en el artículo académico que explica el funcionamiento de la aplicación (Quijano et al., 2018). Además, en una conferencia ofrecida por Lara Quijano (BigData_uc3m, 2022) se explica con un poco más de detalle la labor de los desarrolladores, el origen de la aplicación y por qué se creó. Ello ofrece un poco más de información aparte de la aportada por las instituciones públicas sobre el diseño e implementación de VeriPol.

Respecto a los medios de comunicación, cabe decir que se han hecho eco del uso de VeriPol, tanto en los medios tradicionales como digitales. Se ha de distinguir entre los artículos que informan sobre VeriPol, que repiten una información parecida a las fuentes institucionales [RTVE, 2018; Álvarez, 2019]; artículos que comentan casos resueltos gracias a VeriPol [Europa Press Asturias, 2019; Leonoticias, 2019]; y artículos que opinan y analizan brevemente la actuación de VeriPol y sus servicios (Algorithm Watch, 2020). Especialmente estos últimos se detienen a analizar las implicaciones del uso de la aplicación de la IA, son artículos que se encuentran en sitios web especializados en ciencia y que tratan temas de Inteligencia Artificial. En comparación con los otros dos tipos de artículos, los artículos de prensa digital especializada denotan una investigación

sobre VeriPol más profunda y presentan algunas discrepancias con el funcionamiento de la aplicación.

Se ha de tener en cuenta la limitación de tiempo y espacio. La mayoría de los recursos empleados en la investigación están disponibles de forma abierta en internet. Sin embargo, se ha de recalcar la necesidad de estudiar el caso con testimonios directos de miembros de la policía que hayan usado o no VeriPol, para conseguir una idea más aproximada del funcionamiento diario de la aplicación. Debido a que es una herramienta que incide en la seguridad ciudadana y que se encuentra bajo la tutela de la policía es complicado estudiar la aplicación de la IA, más allá de la información disponible de forma abierta.

No se han podido llevar a cabo las entrevistas a personas que han tenido contacto con VeriPol. Se mandaron correos a los cuatro desarrolladores y se llamó al número de información disponible que ofrecía la policía para pedir información sobre el servicio. Solo dos de los cuatro creadores respondieron, Lara Quijano (Anexo 3) y Miguel Camacho Collados (Anexo 4); y el teléfono de información de la policía comentó que la información es la que está disponible.

6.1.2. Descripción del servicio VeriPol

VeriPol es una aplicación de IA implementada por la Policía Nacional que utiliza una IA, según la UE (Misuraca & van Noordt, 2020, p. 47), relacionada con la robótica cognitiva, el procesamiento automático y conectado. Se emplea a nivel de Administración Local y Central. El proyecto se inició en 2014 (OTRI, 2018) y su propósito es la aplicación de la ley detectando el delito de denuncia falsa. Fue posible su desarrollo gracias a la colaboración de la policía con la universidad, los registros digitales policiales y la exitosa integración con el sistema informático previo de la policía. El impacto esperado de VeriPol es mejorar la detección de denuncias falsas, aumentando así la productividad y reduciendo el número de reportes fraudulentos. VeriPol, como se ha mencionado, es un sistema que usa la Policía Nacional que emplea la IA para detectar denuncias falsas de hurtos y robos, una aplicación que mediante el análisis del lenguaje natural detecta denuncias falsas.

El origen del algoritmo en el que se basa la aplicación es muy sencillo. Miguel Camacho Collados[17], uno de los creadores y policía nacional con formación en estadística y matemáticas, comentó a Lara Quijano, otra de las creadoras del algoritmo, los patrones que observaba en las denuncias falsas (BigData_uc3m, 2022). Observó que parecía haber un mismo *modus operandi* entre quienes hacían las denuncias falsas: una misma estructura de la historia, causas de denuncias falsas parecidas, perfil parecido de las personas que las interponían… (BigData_uc3m, 2022). Por lo que se podía identificar un patrón recurrente que parecía ser siempre el mismo. El delito de "acusación y denuncia falsa y simulación de delito" es común en España, especialmente en los casos de menor gravedad, pero este tipo de delitos puede conllevar unas consecuencias bastante serias, puesto que se incurre en un gasto ingente de recursos valiosos y que además suele

[17] Los desarrolladores de VeriPol, también responsables de otros servicios públicos prestados con IA como VioGén son: Lara Quijano-Sánchez (UAM) Federico Liberatorea (UCM), José Camacho Collados (La Sapienza) y Miguel Camacho Collados (Secretaría de Seguridad, Ministerio del Interior).

acompañarse de un comportamiento fraudulento por parte de los denunciantes que entorpece las labores policiales (Misuraca & Van Noordt, 2020).

Si los reportes falsos siempre seguían un patrón parecido se daba la posibilidad de que pudiera extraerse un algoritmo capaz de definir los casos en los que una denuncia puede considerarse falsa. Gracias a la combinación de la metodología de Procesamiento de Lenguaje Natural (NLP), y los algoritmos de clasificación del *Machine Learning* en el marco de la Ciencia de Datos, VeriPol es capaz de ofrecer una estimación de la probabilidad de que la denuncia sea falsa con una gran precisión (91%) (Quijano et al., 2018).

No es un "detector de mentiras" cualquiera, es un sistema novedoso que no se parece a otros sistemas de detección de noticias falsas, por ejemplo. Es un sistema de IA que permite entender cómo la gente miente a la policía y permite detectar las características de una denuncia de robo o hurto verdadera (Quijano et al., 2018). Analiza las estructuras de las oraciones, el número de adjetivos, el número de verbos, o incluso, si se realizan o no descripciones exactas de los hechos. Dando como resultado que las denuncias falsas suelen ser más cortas, con una descripción menos precisa y suelen abundar los sustantivos (BigData_uc3m, 2022).

Su funcionamiento es muy simple. Una persona acude a interponer una denuncia a la policía. El policía toma declaración al denunciante y registra la denuncia. Después, otro policía conocedor del funcionamiento de VeriPol transcribe la denuncia en el sistema y este le devuelve automáticamente la probabilidad de que la denuncia sea falsa. En ese momento se decide si proceder o no con la investigación del presunto delito – el expuesto en la denuncia o el cometido al presentar la denuncia falsa-. La resolución de VeriPol simplemente ayuda a la toma de decisiones, su nivel de precisión es muy alto, pero se ha de recalcar que siempre ha de ser revisada su estimación por policías expertos en VeriPol, por si hubiera cualquier problema. Además, este sistema se integra a la perfección con SIDENPOL, el sistema electrónico de datos de la Policía Nacional (Dirección General de la Policía, 2021)[18].

En 2015 se realizó el entrenamiento del algoritmo del sistema de IA con denuncias falsas y verdaderas correspondientes a casos reales. Para ello se entrenó al sistema con 1122 denuncias por robo y hurto, de las cuales 588 eran falsas y 534 eran verdaderas, alcanzando un 90% de acierto. Se ha de destacar que dichas denuncias solo incluían el texto de la denuncia, no se tenía en cuenta los datos del denunciante (Dirección General de la Policía , 2018). Una de las características más interesantes de este sistema es que solo necesita el texto de la denuncia para hacer su estimación, y es ese uno de sus puntos fuertes, puesto que evita los posibles sesgos que pudiera haber por raza, sexo o localización (BigData_uc3m, 2022).

Más tarde en junio 2017, se desarrolló una prueba piloto en Málaga y Murcia. En solo una semana se detectaron en Málaga un total de 39 denuncias falsas y en Murcia, 25, cuando el promedio de los meses de junio entre 2008 y 2016 fue en Málaga de 12 denuncias falsas y en Murcia, 3. Comparando ambos períodos 2008-2016 y 2017, la eficacia en la detección de denuncias falsas fue en Málaga de un 84,78% y en Murcia de un 81,58% [Dirección General de la Policía, 2018; Quijano et al. 2018].

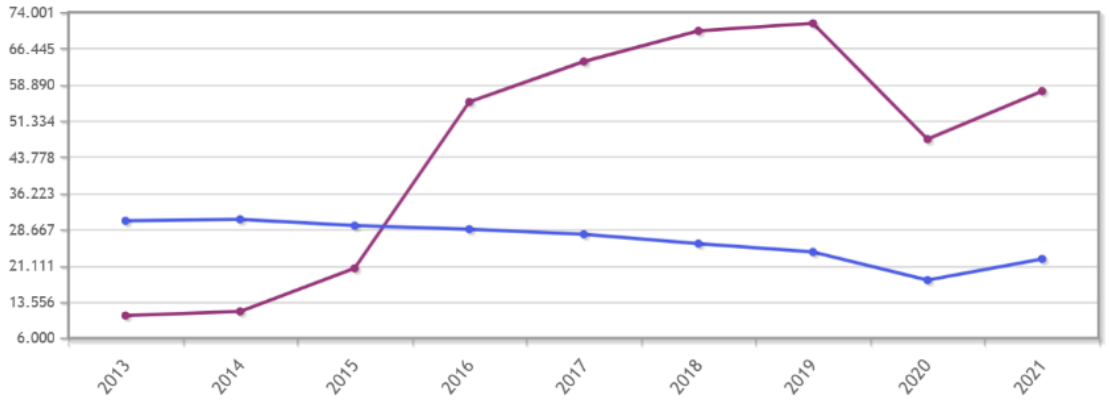
[18] Véase Dirección General de la Policía (2021), se explica como procede la policía con el tratamiento de sus bases de datos: da opción a acceso, rectificación y supresión.

En el mismo período de tiempo, tanto en 2015 como en 2017, un experto analizó las mismas denuncias que analizó VeriPol, pero obtuvo una tasa menor de acierto.

Desde su implementación en octubre del 2018 en las 240 comisarías de policía existentes en España- exceptuando las comunidades de País Vasco, Navarra y Cataluña que tienen sus propios cuerpos de seguridad- se ha empleado en 84.000 casos con una tasa de acierto de aproximadamente el 90% (Algorithm Watch, 2020). Encontramos en prensa algunos casos en los que se aplicó VeriPol, por ejemplo, en Salamanca (Redacción Salmanca24horas, 2019), en Andalucía [R.G., 2019; Noticias de Almería, 2021], en Asturias [Europa Press Asturias, 2019; Europa Press, 2019] y en León (Leonoticias, 2019). Las cuatro noticias mencionadas son noticias muy cortas, en las que se detalla poco lo ocurrido, se menciona que se ha resuelto el caso gracias a VeriPol, pero ninguna comenta cuáles fueron los resultados del análisis del sistema. Solo se hace referencia al funcionamiento de VeriPol en general, pero no se específica como el sistema de IA ayudó en concreto en el caso.

Se señala que ha ayudado a reducir los delitos de hurto y de robo, disuadiendo a la población de cometer tales delitos y por la detección de denuncias falsas de hurto y de robo (Álvarez, 2019). La siguiente gráfica presenta los datos del INE de condenados por hurtos (rosa) y robos (azul). En el caso del delito de robo sí se ve un ligero descenso, pero también ha de tenerse en cuenta que con la pandemia de COVID-19 hubo un brusco descenso de ambos delitos:

Gráfico 1. Estadística de condenados por Hurtos y Robos en adultos (Total)

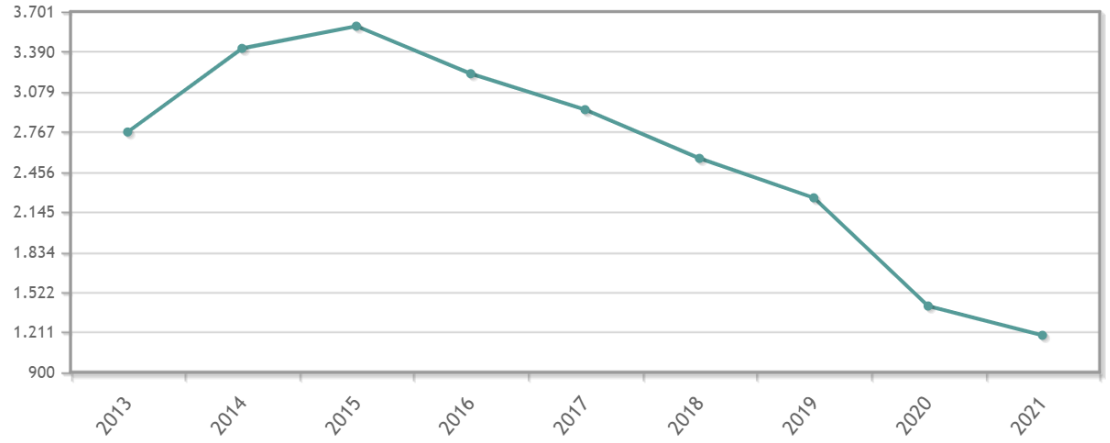


Fuente: INE (2023a)

Por lo que sería complicado establecer una correlación directa entre la implementación de VeriPol y la reducción del número de hurtos y robos.

Respecto al número de condenados por acusación y denuncia falsa y simulación de delito, se observa una tendencia a la baja, que se acentúa en 2019. Sin embargo, cabría hacer un estudio más detallado de la correlación entre estos datos y la implementación de VeriPol, para así comprobar si el descenso de condenados se debe a la disuasión de comisión de delito por el uso VeriPol. Aunque habría de preguntarse si este delito, en vez de bajar sus números, debieran de haber aumentado.

Gráfico 2. Número de condenados acusados por Acusación y Denuncias falsas y Simulación de delitos (Adultos)



Fuente: INE (2023b)

Teniendo presente lo anterior, se ve como las herramientas que emplea la Administración en su prestación de servicios han ido cambiando. En un primer momento tanto el registro, la toma de declaración de los denunciantes y el análisis de las denuncias se hacían de forma manual. En un segundo momento, la electronificación de la administración ha dado lugar, por ejemplo, al registro de datos electrónicos de las denuncias y denunciantes que pasa a ser electrónico (SIDENPOL). Y en un tercer momento, se implementa la smartificación de la administración, se emplean sistemas de IA capaces de analizar y comparar entre todos los registros de denuncias el lenguaje empleado para discernir si la denuncia por la que se le pregunta es verídica o simulada.

Además, se ha de resaltar que esta aplicación no hubiera sido posible sin el perfil multidisciplinar de los desarrolladores y el trabajo que se realizó con los usuarios finales. La colaboración entre quienes elaboran el sistema de IA, así como todos los que participan en el ciclo de vida del sistema, es necesaria para la buena consecución de la smartificación de los servicios públicos.

El servicio VeriPol no solo ayuda a detectar denuncias falsas, sino que de él se pueden extraer las diferentes formas o patrones de los que las personas se sirven para mentir a la policía. La detección de denuncias falsas por robo y hurto es uno de los primeros pasos de este "detector de mentiras" sobre el papel. Los desarrollos del algoritmo de VeriPol actualmente se desconocen, Lara Quijano (BigData_uc3m, 2022) comenta que cedieron el algoritmo al cuerpo de seguridad y que las diversas aplicaciones o funcionalidades que hayan podido desarrollar solo se conoce a nivel interno de la policía, aunque el algoritmo y el artículo académico donde se explica su funcionamiento son públicos.

6.2. Análisis de riesgos del ciudadano

6.2.0. Derechos humanos

El punto de partida para determinar si una IA es fiable y no supone ningún riesgo para los seres humanos en su empleo, especialmente en la prestación de servicios públicos, son los derechos humanos. Los derechos humanos se han de tener en cuenta a la hora de valorarse el riesgo de la IA como un aspecto esencial. Los aspectos a los que se ha de prestar atención son: las potenciales discriminaciones negativas, si se protege a los menores, si se sigue la normativa básica de protección de datos y si respeta la libertad de expresión, información y asociación.

En primer lugar, respecto a la potencial discriminación negativa de VeriPol se comenta que es probable que incurra en discriminaciones negativas. La aplicación está diseñada de tal forma que el propio modelo es capaz de detectar si se están produciendo sesgos por razón de sexo, color, edad, etc. Dichos sesgos se detectan en los pesos que tiene cada variable del algoritmo, pudiendo así comprobar rápidamente si la IA ha desarrollado una serie de sesgos hacia un cierto grupo de población. Además, la información que analiza y tiene en cuenta la aplicación, para detectar si una denuncia es falsa o no, es el cuerpo de texto de la propia denuncia, sin atender a los datos del denunciante, hecho que pudiera generar sesgos (BigData_uc3m, 2022).

Ante esta cuestión se ha de señalar que en el mismo reporte de la denuncia pueden darse datos concretos sobre raza, color, sexo y edad de las personas que han perpetrado el robo o hurto. Porque si bien VeriPol no tiene en cuenta los datos del denunciante y no propicia una tendencia a que ciudadanos con unas características determinadas sea más probable que mientan en sus reportes, si puede haber cierto sesgo en el análisis en el texto de la denuncia. Por lo que comentan los desarrolladores, los sesgos se solventarían con las alertas que se establecerían si los pesos de ciertas variables del algoritmo se encuentran en un nivel alarmante.

En segundo lugar, los derechos de la infancia se respetarían en tanto que VeriPol se desarrolla bajo la normativa europea. Sin embargo, esta afirmación sería una suposición derivada de los otros puntos de la investigación puesto que no se hace una mención explícita a los menores. En el INE se encuentra desglosado el número de condenas de robo, hurto y simulación de delito y falsa acusación en mayores y menores de edad, pero cuando se habla sobre el funcionamiento de VeriPol, nada se menciona respecto a la infancia y la protección de sus derechos. Lo único que llega a mencionarse respecto a los menores, es que una de las denuncias falsas más habituales, en la que un menor interviene, son aquellas en las que acuden junto con sus progenitores a comisaría a interponer una denuncia por el robo de su móvil. Situación que suele darse con móviles de alta gama para el cobro de seguro por la pérdida del móvil del menor (BigData_uc3m, 2022).

En tercer lugar, el sistema VeriPol sigue el Reglamento Europeo de Protección de Datos. Cualquier actualización en la normativa europea y nacional se lleva a cabo también en la aplicación de IA. Se presta especial atención a la protección de datos puesto que se trata con datos personales de alta sensibilidad de los ciudadanos, especialmente por el estrecho nexo que mantienen VeriPol y SIDENPOL. Se reconoce al ciudadano el derecho

a la modificación o rectificación de sus datos, el derecho al acceso de su información policial y el derecho a la supresión de sus datos según lo establecido en el art. 18.4 de la Constitución Española, en el Registro de Actividades de Tratamiento del Ministerio del Interior, en la Ley Orgánica 3/2018 de Protección de Datos Personales y garantía de los derechos digitales y en el Reglamento General de Protección de Datos [Quijano et al., 2018; Dirección General de la Policía, 2018].

En cuarto lugar, no se mencionan de forma explícita posibles interferencias en el respeto a la libertad de expresión, información y libertad de asociación. En un principio no tendría que suponer ningún riesgo ni amenaza.

6.2.1. Intervención y supervisión humanas

A la hora de la toma de decisiones, en este caso, es el policía quien determina si la denuncia es falsa y la incurrencia en delito de la persona que la interpone, pero su juicio puede verse influido por la IA más de lo esperado. La IA supone un riesgo para el ciudadano en el momento en el que la decisión del sistema informático prima sobre la decisión humana en dicha cuestión. En el caso a estudiar, la actuación del policía que conlleva que la IA se convierta en un potencial riesgo, es la plena confianza y nula revisión de los resultados del análisis de la denuncia que realiza VeriPol. Por ello, se ha de tener en cuenta lo siguiente: hasta qué punto el sistema de IA puede interactuar con los policías o guiarlos en su toma de decisiones, si los resultados de su análisis generan confusiones y cómo ello afecta a la toma de decisiones autónomas, y si el sistema de IA interactúa con los usuarios finales.

En primer lugar, VeriPol no está diseñado para interactuar con los policías a la hora de afirmar que una denuncia es falsa o no. VeriPol hace una estimación de la probabilidad de que la denuncia sea falsa, solo guía u ofrece información que puede ser relevante a la hora de decidir, por parte de los policías, si la denuncia en cuestión es verdadera o falsa. Es un soporte, pero nunca ha de ser "quien" tome la decisión, hecho que recalcan los desarrolladores (Quijano et al., 2018).

En segundo lugar, no debieran de generar confusiones los resultados emitidos por VeriPol. Se ha diseñado para que sea lo más claro posible, de ahí que su funcionamiento sea automático. Asimismo, la trazabilidad del sistema, según cuentan los desarrolladores, es muy clara precisamente para evitar las confusiones (Quijano et al., 2018). Se ha de tener en cuenta que es posible que el algoritmo pueda equivocarse en su predicción o se dé el caso que pueda reportar información confusa. Hay que contemplar esa posibilidad, pero se ha diseñado VeriPol de forma que las confusiones sean las mínimas posibles.

En teoría no debería de afectar a la toma de decisiones, puesto que es una herramienta con altos niveles de eficiencia y efectividad, pero se ha de volver a recalcar que la decisión final sobre si la denuncia es falsa o no depende del policía encargado de procesar la denuncia. No interactúa como tal con los usuarios finales, solo vuelca el análisis realizado de la denuncia directamente al policía encargado de analizar la denuncia con VeriPol.

En tercer lugar, no habría riesgo de adicción a la IA o manipulación por parte de este sistema de IA. Es una IA débil por lo que no debería de haber riesgo de manipulación. En caso de que esto sucediera o que hubiera algún error en el funcionamiento del sistema que derive en adicción, o que la IA pueda manipular al usuario final, se han establecido

estrategias para evitarlo por parte de los desarrolladores y seguramente por parte de la policía. Aunque no son conocidas dichas estrategias, especialmente las desarrolladas por la policía.

Sin embargo, si bien no hay riesgo de manipulación puede ocurrir que, al ser una herramienta con un alto nivel de precisión y confianza, el experto encargado del uso del sistema no revise los resultados de los análisis emitidos por VeriPol. El experto da por sentado que el análisis que reporta VeriPol es 100% seguro dado sus altos niveles de fiabilidad. Por ello siempre ha de tenerse presente un margen de error, aunque este sea pequeño.

VeriPol, al igual que todo sistema de IA ha de ser supervisado por un humano. No se puede realizar una toma de decisión rápida sin revisar los resultados de los análisis reportados por la IA. Porque tal y como resalta Francisco Álvarez en La Vanguardia (Álvarez, 2019), responsable operativo de VeriPol: "La policía está obligada a investigarlo todo y VeriPol nunca va a decidir, pero nos quita muchísimo trabajo, eso es innegable".

6.2.2. Robustez y seguridad

Es importante que el sistema de IA, VeriPol, demuestre ser robusto y seguro, sobre todo por tratar con datos personales pertenecientes a denuncias, lo cual exige una mayor rigurosidad en el tratamiento de la información. Por ello se han de detectar cuales son los puntos críticos de la aplicación y las posibles estrategias a implementar en caso de que haya cualquier ataque al sistema. Además, se ha de tener en cuenta si el funcionamiento de VeriPol es estable, lo que la dotaría de una mayor seguridad. Por ello, ha de tener mecanismos de evaluación y monitoreo continuo a lo largo de su vida.

En primer lugar, por lo que se ha podido observar, no se comenta nada sobre las diferentes estrategias que pudieran haberse desarrollado para evitar o enfrentarse a posibles hackeos o ciberataques, aunque sí se menciona la necesidad de que las bases de datos de VeriPol estén actualizadas para evitar cualquier problema de ataques. Un sistema informático actualizado es una de las mejores prevenciones ante los posibles riesgos de ataques. Los desarrolladores presentan la aplicación de IA como una aplicación robusta.

Se ha informado a los usuarios finales, la policía, de la necesidad de mantener el sistema actualizado. Si bien no se sabe a ciencia cierta si se han continuado con los procesos de actualización, puesto que se corresponde con información interna de la policía. Se supone que han seguido con los procesos de actualizaciones correspondientes y los análisis de posibles amenazas. También se ha analizado la posible dependencia del sistema de inteligencia artificial y recalcan que es siempre el policía el que tiene la última palabra. Pero tampoco son muy claros a este respecto, ya que podría darse la situación que se ha comentado anteriormente en la que el policía no realiza la comprobación correspondiente del sistema y asume la respuesta de VeriPol como verdadera instantáneamente.

Dada la efectividad de VeriPol podría asumirse su análisis como correcto y no revisar sus procedimientos, lo que supone un riesgo para el denunciante y el ciudadano puesto que se deja en manos de una máquina la decisión de si su testimonio es falso o verdadero en base a un análisis predictivo del lenguaje.

En segundo lugar, gracias al *machine learning* es capaz de irse adaptando y de ir aprendiendo con cada análisis que realiza, lo que asegura que su comportamiento sea estable. Los desarrolladores comentan que tal y como VeriPol está diseñada su comportamiento es estable, pero faltaría la confirmación de los usuarios finales para saber si la aplicación muestra un comportamiento estable. Información que es de carácter interno de la policía, por lo que no es accesible al público general. Sin embargo, si VeriPol se hubiera descontrolado, quizás los medios de comunicación se hubieran hecho eco y hubiéramos sabido que el comportamiento de esta herramienta ha dejado de ser estable.

Uno de los problemas a los que puede enfrentarse VeriPol es un declive en sus niveles de precisión. No se explicita en ningún momento qué sucede si esta situación se da. Solo los desarrolladores recalcan la necesidad de mantenerlo actualizado y realizar revisiones periódicas (Quijano et. Al, 2018). Quizás la policía tenga establecido una serie de protocolos para estos casos, pero esa información no está disponible. En general, si se da una bajada de precisión de VeriPol, ello se muestra en los pesos de las variables. Se puede detectar en qué variable del algoritmo hay un problema de precisión, hecho que también puede ser detectado fácilmente si se mantiene un monitoreo regular de VeriPol.

Por ello, como comentan los desarrolladores, se ha de perseguir que las bases de datos siempre estén actualizadas para evitar los problemas de precisión y de representatividad del entorno. Por ejemplo, si hay cambios en el uso de ciertas formas verbales y gramaticales el sistema ha de ser capaz de detectarlos. Que las bases no estén actualizadas o que el algoritmo no se adapte o aprenda con cada caso que analiza es una cuestión que preocupaba a los desarrolladores, por ello se esforzaron en diseñar el algoritmo de tal forma que se detectaran los problemas de actualizaciones, de precisión y de sesgo. Pero no hay información actual de cómo ha evolucionado el algoritmo porque es una información que maneja la policía de forma interna. Se supone que continúa su monitoreo, que tiene sus bases actualizadas y hay establecidos planes de contingencia en caso de que VeriPol cometa un error.

En tercer lugar, los desarrolladores comentan que se ha proporcionado información y formación específica a los policías para usar y afrontar los problemas que VeriPol pudiera causar. Los creadores comentan que precisamente se trabajó con ellos mano a mano para así establecer los riesgos mínimos y la máxima eficacia posible. Los policías han participado en su proceso de elaboración y comentado los errores y riesgos que veían en la implementación de VeriPol. Los creadores afirman que fue gracias al equipo multidisciplinar que se estableció, que se facilitó la detección de posibles riesgos, ya que los policías señalaban aspectos, que a lo mejor ellos no hubieran considerado como importantes al no estar a diario en la gestión de denuncias de la comisaría.

Sin embargo, hay quien considera que no se hizo lo suficiente por una implantación adecuada. En 2020, en Vigo llevaban dos años sin utilizar VeriPol, situación que se agravó con la pandemia del COVID-19, según fuentes autorizadas de la comisaría de Vigo (Pita, 2020)[19]. VeriPol, por aquellas fechas, ya estaba implementada según fuentes oficiales, en todas las comisarías, pero al parecer la formación para saber cómo utilizarla en las comisarías se realizó de forma asimétrica, ya que en la comisaría de Vigo no habían recibido dicha formación. A ello se añade lo que comenta en la misma noticia

[19] "La versión que dan fuentes autorizadas de la comisaría de Vigo es que la aplicación no la usan todavía porque al no haber sido instalada no se ha dado la formación al personal que va a emplearla y nadie sabe cómo funciona. Pero desde la Jefatura Superior de Galicia y desde la Dirección General de Policía sorprende esta respuesta, ya que VeriPol funciona en red en todas las comisarías de España desde eficacia en A Coruña, Barcelona, Murcia, Valencia, Sevilla y otros centros policiales" (Pita, 20/08/2020).

de La Voz de Galicia el portavoz del sindicato de la Unión Federal de Policía, Agustín Vigo, que afirma que VeriPol no se utiliza en bastantes comisarías porque el programa no funciona adecuadamente. Apunta que algo debió pasar para que funcionara de forma tan precaria, a lo que se sumó la poca formación de los policías y la pandemia del COVID-19.

Se ha de señalar que cuanto menor sea la información disponible sobre este tipo de programas que incorporan IA y mayor desconocimiento haya sobre su funcionamiento es posible que haya más rechazo a su uso. Puede verse como algo peligroso, pensar que el programa supondrá un gran cambio en la organización y en las formas de trabajo. Se ha de tener presente que aún estamos lejos de alcanzar la transformación completa de las organizaciones por el uso de la IA en las tareas que se llevan a cabo en estas. De ahí la importancia de que la implementación de la IA sea igual para todo el territorio. En la implementación de VeriPol pudo haber un problema en la zona de Vigo, y si de por sí la IA puede generar inseguridad, si en el proceso de implementación hay problemas, la desconfianza aumenta.

6.2.3. Privacidad y gestión de datos

Una de las mejores maneras de asegurar un programa informático que emplee IA y sea robusto y seguro es plantear una buena gestión de datos y proteger la privacidad de los ciudadanos, por ello ha de examinarse si en su elaboración y desarrollo se han seguido las normativas correspondientes, qué tipo de datos se empleó en el entrenamiento del algoritmo y si se puede garantizar la privacidad de los datos personales que se emplean en estas aplicaciones, así como la gobernanza y gestión de los datos, según lo establecido en las normativas correspondientes.

En primer lugar, se ha seguido tanto en el procedimiento de desarrollo como en su implementación la normativa de regulación de datos, según afirman los creadores de VeriPol, a nivel europeo (Reglamento Europeo de Protección de Datos o GDPR)[20] y estatal (Ley Orgánica 3/2018 de Protección de Datos Personales y garantía de los derechos digitales o LOPDGDD). La actual gestión que realiza la policía también seguiría esos cauces, aunque debería de corroborarse con fuentes policiales.

Sin embargo, en un primer acercamiento, a la hora de interponer una denuncia se observa que no se informa al ciudadano de que su denuncia puede ser revisada por VeriPol. Habría que determinar en un análisis más detallado si esta cuestión debiera de ser revisada para que se incorporara en la información previa que se da al denunciante cuando presenta la denuncia. Los datos que aporta en su denuncia pueden ser analizados con un sistema de IA para comprobar si es verdadera o falsa, y quizás ello debiera aparecer en el documento de información de los derechos a la víctima del delito que se entrega al hacer la denuncia, aunque es posible que no se informe de ello porque puede que no se utilice para analizar la veracidad de su denuncia o porque todavía solo se aplica a delitos de hurto o robo. Pero cabe preguntarse si debiera de informarse a la supuesta víctima, independientemente de si se utiliza la aplicación o no.

[20] También habría de tenerse en cuenta la Directiva 2002/58/CE sobre privacidad y comunicaciones electrónicas.

En segundo lugar, una cuestión que sí que se recalca respecto a la gobernanza y gestión de datos es que la decisión de que el equipo fuera multidisciplinar fue a propósito para esbozar una mejor gobernanza y gestión de datos, aunque ahora la gestión dependa de la policía. A lo que se suma que el entrenamiento que se realizó en 2015 se hizo con denuncias reales, por lo que ya en las primeras pruebas se trabajó con contenido y datos personales de los ciudadanos.

Aunque no se hace mención explícita al modelo de gobernanza de datos que se emplearía, se intuye que sería el modelo *goverment to goverment*. El tipo de datos que se comparten son datos de titularidad pública por norma general, sobre todo porque los delitos a los que de momento se aplica VeriPol- robos y hurtos-, no se suelen requerir datos de procedencia privada. Excepto, quizás, si el caso se relaciona con el cobro de un determinado seguro en el que sea necesario que la entidad aseguradora y la policía compartan datos, por lo que estaríamos ante los modelos G2B o B2G.

Por ello los protocolos de protección de datos son tan importantes, y que tanto los creadores como quienes emplean diariamente la aplicación sean capaces de asegurar la integridad de los datos que procesan. Esta cuestión se menciona de pasada tanto en el artículo académico de los creadores (Quijano et al. 2018) como en la charla de Lara Quijano (BigData_uc3m, 2022) sobre VeriPol. Además, otro comentario que señalaban los desarrolladores en diversas entrevistas que se les han realizado, tanto en el periódico (Álvarez, 2019) como televisión (RTVE, 2018), era que sí tuvieron en cuenta la protección de la privacidad y la defensa de una buena gestión de datos.

Los desarrolladores recalcaban la buena gestión de datos precisamente por el nivel de sensibilidad de los datos que VeriPol empleaba. Aunque VeriPol solo analiza el cuerpo de texto de la denuncia y no los datos del denunciante, no dejan de ser datos altamente confidenciales y la aplicación está conectada con SIDENPOL (Dirección General de la Policía, 2021). Por lo que la gestión y gobernanza de los datos ha de realizarse con la máxima seguridad.

Una de las preguntas que pueden surgir al respecto es si VeriPol es capaz de guardar datos más allá de los necesarios para llevar a cabo su proceso de aprendizaje. Es decir, si aparte de guardar las estructuras sintácticas del lenguaje y el estilo de registro empleado, guarda las localizaciones que se mencionan en la denuncia, aunque estas se omitan a la hora de determinar si la denuncia es falsa o no. Porque hasta qué punto el algoritmo elimina y no recuerda los aspectos de la denuncia que pueden generar sesgos.

En el caso de que se guardaran datos de los ciudadanos, no del denunciante sino de aquellos que aparecen en el cuerpo de texto de la denuncia, y se quedaran guardados en SIDENPOL (Dirección General de la Policía, 2021), los ciudadanos deberían poder solicitar el acceso para ver sus datos en esta plataforma, así como modificar o pedir la supresión de sus datos personales de dicha base. Para realizar estas acciones deberán de proceder mediante una solicitud al órgano correspondiente tal y como se especifica en la página informativa de la Dirección General de Policía (2021).

Otro de los posibles problemas a la hora de administrar y gestionar los datos que el denunciante proporciona a la policía es la transcripción que el policía realiza. El artículo informativo de AlgorithmWatch (2020) sobre VeriPol y las opiniones que realizan algunos ciudadanos en Telemadrid (Sindicato CPPM, 2018) al ser preguntados por esta aplicación de IA, apuntan una serie de dudas que han de ser tenidas en cuenta en relación con la transcripción que el policía realiza de la narración del denunciante.

VeriPol analiza lo que el policía encargado y experto en el funcionamiento de VeriPol introduce en el programa, pero el policía encargado de la transcripción puede hacer una transcripción inexacta de lo que el denunciante narra, por lo que el resultado del análisis del algoritmo puede no ser del todo exacto, porque no ha introducido de forma exacta las palabras utilizadas por el demandante (AlgorithWatch,2020). A ello se suma, que las personas que acuden a denunciar pueden tener dificultades a la hora de expresarse por lo que el algoritmo, al no entender los matices de un uso más común del lenguaje puede ofrecer un análisis erróneo (Sindicato CPPM, 2018).

Estas cuestiones no se comentan en el artículo académico (Quijano et al. 2018) ni en la charla ofrecida por Lara Quijano sobre VeriPol (BigData_uc3m, 2022). Pero una de las respuestas que quizá se ofrecieran es que, poco a poco, el algoritmo irá aprendiendo los usos del lenguaje y el policía habrá de ser muy cuidadoso con como transcribe el reporte de la denuncia en el programa informático para que se siga el procedimiento adecuado y no haya ninguna incidencia. Ambas situaciones pueden poner en duda la neutralidad de VeriPol, porque el problema no se encontraría en el análisis que realiza el algoritmo, sino en la información que se le proporciona. La información estaría sesgada o habría recibido un "tratamiento" previo, de ahí la importancia de una buena gestión de los datos.

Los desarrolladores enfatizan el bajo peligro que supone para la privacidad y la gestión de datos personales VeriPol, puesto que lo importante del texto, y lo que analiza el sistema de IA, es aquello que se narra, el acontecimiento, no los rasgos de los atacantes o la localización determinada del hurto o robo, ya que esta información sí que podría generar sesgos. En un principio se evitan los datos personales del denunciante como el nombre, municipio, edad, etc. Pero la pregunta es hasta qué punto VeriPol es capaz de hacer caso omiso a esos detalles si estos se encuentran en el cuerpo de la denuncia. A lo que se suma, que no se informa al ciudadano sobre la utilización de esta herramienta cuando interpone una denuncia y si ello puede suponer una vulneración del derecho a la información de los ciudadanos sobre cómo van a ser tratados sus datos.

Estas cuestiones son especialmente relevantes sobre todo porque VeriPol servirá como base para el desarrollo de otras futuras aplicaciones en el seno de la policía. Aplicaciones que estarán ligadas con la creación de perfiles (BigData_uc3m, 2022) y que la Interpol seguirá de cerca, puesto que también estaba muy interesada en ver como se desarrollaba el proyecto (France 24 Español, 2018).

6.2.4. Transparencia

La transparencia es uno de los ejes esenciales para minimizar los riesgos que pueda presentar una IA en la prestación de servicios públicos. La trazabilidad de los pasos que se siguen tanto en su diseño como en su implementación han de ser rastreables para una mayor transparencia y seguridad. Así como la opinión y consideraciones de los usuarios finales, los policías, que muestran si se les ha informado adecuadamente sobre el funcionamiento y límites de la herramienta.

En primer lugar, los desarrolladores afirman que se puede realizar una trazabilidad clara del algoritmo. El algoritmo se encuentra disponible en abierto y cualquier mejora que se realice por parte de los desarrolladores también se hará pública; además, están abiertos a que cualquier ciudadano pueda proponer una mejora del algoritmo. Ahora el

algoritmo es gestionado por la policía por lo que son ellos quienes deben de presentar la rendición de cuentas. Las mejoras que se implementen al algoritmo por parte del ámbito policial solo serán conocidas a nivel interno de la organización (BigData_uc3m, 2022). Este hecho manifiesta una pérdida de transparencia en comparación con la actitud de los creadores de ir añadiendo y actualizando las modificaciones del algoritmo.

En segundo lugar, como se ha comentado tanto desde fuentes policiales (Pita, 2020) como en la charla de Lara Quijano (BigData_uc3m, 2022), se indica que ha habido un contacto continuo con los policías que han participado en el proceso de elaboración del programa. Ello ha permitido una mejor adaptación del algoritmo al servicio público que posteriormente prestaría la aplicación de IA. Que se haya trabajado de forma cercana con los usuarios finales sin duda alguna ha sido muy beneficioso y han conocido el funcionamiento del programa de primera mano, pero quizá no todos los usuarios conozcan cómo funciona VeriPol.

Los desarrolladores comentan que sí se ha informado sobre el funcionamiento del sistema y de sus limitaciones. Pero como se ha mencionado en anteriores categorías de análisis, no todos los policías se mostraban satisfechos con la implementación de VeriPol. no conocían su funcionamiento, no habían recibido la formación adecuada y aunque el sistema estuviera disponible en sus comisarías, no podían utilizarlo (Pita, 2020). Cierto es que estas informaciones corresponden a 2020, por lo que la situación podría haber cambiado y ya haber un conocimiento del funcionamiento de VeriPol en todas las comisarías. Sin embargo, este hecho muestra lo importante que es recibir la formación para poder utilizar el sistema y evidencia que el proceso de formación ha sido desigual en los diferentes territorios españoles.

Se ha de comentar que, a la hora de hablar sobre transparencia, no solo se ha de tener en cuenta si el algoritmo es público o la consideración de los usuarios finales o *stakeholders*, esenciales para llevar a cabo una buena integración de la IA en la Administración y sus servicios públicos, sino que también ha de tenerse en cuenta la accesibilidad de la información sobre la aplicación. La información pública disponible que se ha encontrado sobre VeriPol es la esencial, las características básicas del algoritmo y sus pruebas piloto. También se encuentra información sobre algunos casos en los que ha intervenido en su resolución VeriPol, pero no se comenta qué es lo que resultó clave para resolver el caso, ni se comenta el análisis que realizó el programa para determinar la probabilidad de falsedad de la denuncia.

Se afirma que todos los datos acerca de VeriPol son públicos, pero tal y como señala AlgorithmWatch (2020) no hay disponible información exhaustiva ni estadísticas sobre la aplicación. En el INE solo se puede consultar el número de delitos cometidos de robo, hurto o acusación y denuncia falsa, pero no hay una relación de datos sobre VeriPol como si los hay de VioGén (Dirección General de Coordinación y Estudios, 2023)[21], por ejemplo. No se encuentra el número de denuncias falsas detectadas por VeriPol, solo si acaso en algún periódico provincial en el que se ha resuelto un caso gracias a VeriPol; ya que quien redacta la noticia especifica el número de denuncias falsas detectadas por VeriPol en la comisaría, datos que han sido confirmados por su fuente policial, como los casos que se han referido al inicio de este apartado.

Cierto es que VeriPol es un programa utilizado por la policía para detectar denuncias falsas, por lo que la seguridad juega un papel clave en el caso a analizar. Seguramente se prima la seguridad al derecho a la información, para tener un mayor

[21] Véase (Dirección General de Coordinación y Estudios, 2023).

control sobre la herramienta de detección de denuncias falsas, pero esta suposición necesitaría la confirmación por parte de fuentes policiales. Aunque, este razonamiento no es del todo descabellado, porque así los potenciales delincuentes solo disponen de una información pública básica, que los informa a nivel general de VeriPol, pero no poseen información específica sobre cómo es su funcionamiento cotidiano para que así no busquen formas de sortear su análisis, como deja entrever Lara Quijano en su charla (BigData_uc3m, 2022). Especialmente, cuando la policía tiene como objetivo la aplicación de VeriPol en casos más complejos.

Aunque sea necesario cierto control sobre la información pública de VeriPol, puede que con el tiempo haya más datos disponibles. Sin embargo, aunque estos no sean accesibles para el público en general, si debieran de serlo para los investigadores ya que VeriPol es un sistema de IA novedoso tanto para el sector policial como para el ámbito académico – ingenieros, politólogos, sociólogos, etc. -.

6.2.5. Diversidad, no discriminación y equidad

A la hora de valorar el riesgo que puede suponer para el ciudadano o el usuario final los sistemas de IA, se ha de tener en cuenta la diversidad de los usuarios a los que va dirigida, así como la diversidad de quienes han de utilizarla para no generar situaciones de discriminación negativa e inequitativas. En el caso de VeriPol, quien emplea y es destinatario final del programa informático son los propios policías, por lo que habrá de estudiarse si hay alguna barrera que impida a algún miembro de la policía acceder al sistema de VeriPol. Asimismo, se ha de tener en cuenta los sesgos que pueda generar en los análisis de las denuncias el algoritmo de VeriPol. En ambos casos para prevenir posibles riesgos para el ciudadano se debe de haber establecido una estrategia para la minimización de discriminación y sesgos, así como su medición o la definición correcta de las variables que integran el algoritmo.

En primer lugar, se ha de comentar si se han considerado las dificultades que los policías pudieran tener a la hora de utilizar VeriPol, así como los sesgos que la aplicación pudiera generar a consecuencia de sus análisis. Gracias a la colaboración en el desarrollo que se realizó entre los creadores y los policías, se fue adaptando la aplicación según los comentarios que los futuros usuarios finales realizaban. Por ello se afirma que sí se tuvo en cuenta las posibles dificultades de los policías porque se les escuchó durante el proceso. Sin embargo, como se ha comentado líneas más arriba, uno de los problemas para poder entender bien el funcionamiento de la aplicación es la formación, porque sin ella no se sabe cómo emplearla adecuadamente, por lo tanto, no se usa y puede generar desconfianza entre el equipo policial (Pita, 2020). A ello se añade que solo aquellos formados en VeriPol pueden manejar el sistema. Al parecer, no todos los policías tendrían acceso al sistema. Esta información sobre el funcionamiento diario de VeriPol y si hay en concreto un departamento o un grupo de policías específico para el empleo de VeriPol, serían fuentes policiales quienes habrían de confirmarlo.

Por otro lado, también pueden darse sesgos en el análisis que realice VeriPol de las denuncias. El algoritmo, con cada análisis, registra los datos, las características y las palabras clave que le han llevado a considerar la denuncia verdadera o falsa y puede detectar que cierto tipo de información sobre raza, edad o sexo se repite más frecuentemente que otra. Rastrear estos sesgos es posible gracias a las ponderaciones de las variables, que ya se han comentado. Su monitorización hace que se observen sus

niveles de impacto y así ver cuál es la variable que necesita reajustarse. Por ejemplo, en las pruebas piloto que realizaron se dieron cuenta de que si reducían los niveles de eficiencia y precisión del sistema – de un 90% a un 80% aprox.-, se reducía la discriminación negativa que realizaba el sistema hacia cierto tipo de población (BigData_uc3m, 2022).

Curiosamente, las denuncias falsas son las que presentan más sesgos, dice Liberatore en una entrevista para RTVE (RTVE, 2018). El investigador italiano comenta que VeriPol no es racista, pero que en las denuncias se encuentran muchos sesgos discriminatorios. Aunque se eliminen los datos personales y concretos de las denuncias, el algoritmo puede tender a reportar comportamientos sesgados. Por ello es muy importante su monitoreo, control y reajuste de las variables para mantener un comportamiento fiable y seguro del algoritmo empleado por VeriPol.

En segundo lugar, y en relación con lo anterior, no solo se han de controlar los pesos de las variables para detectar posibles riesgos y mantener actualizadas las bases de datos, también se han de mantener actualizados los conceptos que emplea la IA para analizar las denuncias. Con el paso del tiempo los conceptos de referencia que emplea el sistema pueden cambiar y se han de actualizar, especialmente el lenguaje coloquial. Esta cuestión también la han tenido presente los creadores y de ahí su insistencia en la realización de monitoreos continuos. Pero, vuelve a ocurrir lo que se ha comentado en otros casos, que ahora dichos monitoreos se realizarían por parte de la policía.

En tercer lugar, respecto a la participación en la elaboración del sistema y los riesgos que supone que no todos puedan acceder a VeriPol, Lara Quijano (BigData_uc3m, 2022) comenta que sí se han tenido en cuenta para hacerla lo más accesible y sencilla posible, sobre todo porque así es mucho más fácil su uso. Sin embargo, se ha de destacar que no todos los policías tienen acceso a este sistema, sino solo aquellos que son expertos o han sido formados en el uso de la plataforma.

Todo parece indicar que la formación sobre la aplicación la ofrecían a todo el cuerpo, pero al ser diferentes policías quienes toman el reporte en un inicio y otros quienes "lidian" con VeriPol, podría implicar que, al final, hubiera una división entre aquellos policías que sí saben manejar VeriPol y de forma continuada y otros que no. Esto podría suponer un problema, porque es importante que el texto que se introduzca en el programa ha de ser más o menos entendible y puede que la transcripción de la denuncia no reporte la información exacta del denunciante lo que supone un tratamiento previo de la información que puede derivar en información sesgada (Sindicato CPPM, 2018).

6.2.6. Bienestar social y medioambiental

Otro de los aspectos a considerar para valorar los posibles riesgos que la IA supone para los ciudadanos es el de la protección del entorno social y natural. Pero, en el caso de VeriPol, no se comenta ningún riesgo medioambiental que pueda presentar su implementación, tanto a nivel de construcción del equipo necesario para hacer posible el análisis como por las consecuencias medioambientales que pudieran generar sus análisis.

Sin embargo, sí se encuentran referencias al impacto que tendrá esta tecnología en la forma en la que se realiza el trabajo humano. En la noticia a la que se ha hecho referencia de La Voz de Galicia (Pita, 2020), los policías a los que entrevistaron comentaban que no era una herramienta de fiar o que iba a suponer la ruina del trabajo

policial, porque el "olfato" que tiene un policía no lo puede tener una máquina. Indudablemente la IA tendrá un gran impacto en las formas de trabajar tal y como comenta Ramió (2019), es una herramienta que ha de ayudar a la realización de ciertos procesos, y aquellos que sean automáticos y repetitivos seguramente acaben siendo desempeñados por una IA.

Pero VeriPol no pretende sustituir el trabajo de los agentes policiales, sino que pretende facilitar y ayudar en la labor de detección de denuncias falsas. El agente es quien finalmente ha de tomar la decisión de si dicha denuncia es falsa o no. Además, para la aprobación de utilización del sistema han contado con la aprobación de los organismos correspondientes. VeriPol nace con el propósito de reducir la carga de trabajo a los policías para poder dedicar mayores recursos a casos que son reales y reportan un verdadero peligro.

Pero, como se ha comentado en anteriores categorías, aunque sí que los desarrolladores afirman que se ha proporcionado formación a los policías para saber cómo emplear el programa informático, parece que ha habido algunas desigualdades en su implementación. Tanto las fuentes del Ministerio del Interior de AlgorithmWatch (2020) como la asociación Policía del siglo XXI recalcan que aun habiendo recibido formación encontraban la aplicación difícil de manejar.

Aparte de lo comentado, se ha de tener en cuenta si VeriPol puede impactar de una forma determinante en la democracia. El empleo de nuevas tecnologías y formas de prestar los servicios públicos también tiene un impacto en los valores y prácticas de las organizaciones públicas, ya que pueden transformar por completo sus estructuras de funcionamiento y de relación con el entorno interno y externo de la institución.

No es un aspecto que comenten sus creadores y el nivel de desarrollo actual en el que se encuentra VeriPol no tendría por qué tener un impacto relevante en la democracia. Pero no se ha de perder de vista la evolución que puede tener esta aplicación, ya que los creadores comentan que se podría seguir desarrollando el algoritmo e implementarlo en casos más graves en los que el análisis de testimonios sea clave para determinar su veracidad o falsedad, puesto que sus niveles de eficiencia eran muy altos.

Dentro de unos años el desarrollo de VeriPol puede tener un importante papel en la seguridad nacional. La policía nacional, que es quien actualmente gestiona el algoritmo, son quienes podrían responder con mayor exactitud a este tipo de preguntas. Se puede, sin embargo, conjeturar sobre los prototipos de perfiles más probables de cometer un delito que la IA puede identificar o pronosticar. Cabría preguntarse el efecto que podría tener en una democracia, puesto que los agentes de seguridad tendrían más controlados a las personas o lugares que coincidieran con ese tipo de características.

Pero, hasta qué punto la cuestión mencionada no genera un determinado sesgo hacia cierto tipo de lugares o personas, no por parte de la máquina sino por parte de los propios ciudadanos. Es decir, a consecuencia de los análisis de las máquinas las personas de una sociedad pueden discriminar a otras personas o ciertos lugares. Sin duda es beneficioso que se puedan predecir los ambientes que puedan resultar más conflictivos, ayuda a focalizar los recursos, pero se ha de tener en cuenta que se habría de establecer una serie de mecanismos claros y transparentes para detectarlo.

Lara Quijano (BigData_uc3m, 2022) comenta que VeriPol no es la tecnología que se presenta en *Minority Report*, pero sus posibles desarrollos sí que pueden recordarnos a la película. Se ha de tener presente que puede tener un potencial impacto negativo si no

se realizan los monitoreos y controles adecuados, sobre todo porque concierne temas de seguridad. Pero para poder establecer un análisis más concluyente habremos de estar pendientes de su evolución en los próximos años dado lo joven que es el campo de estudio y la aplicación de VeriPol.

6.2.7. Rendición de cuentas

La rendición de cuentas es uno de los puntos más importantes, sino el que más, a la hora de evaluar los riesgos que se pueden presentar en los servicios públicos prestados con IA. Es necesaria la realización de auditorías internas y externas, un protocolo propio de evaluación del sistema, la garantía de la protección de los derechos fundamentales, una revisión ética y realizar un monitoreo continuo de los otros puntos comentados.

Comentan los desarrolladores que es posible realizar auditorías externas e internas, pero no se ha podido acceder a información sobre estas auditorías porque esta información solo está disponible a nivel interno en la policía. Aparte de la realización de auditorías, el propio sistema tiene un protocolo de evaluación que se basa en el estudio de las ponderaciones de las variables, que anteriormente se ha comentado, y se pueden establecer alertas para el caso de que ciertas variables presenten valores muy altos. Pero si hay algún otro sistema de evaluación no hay información disponible sobre ello.

Por otro lado, se comenta que los derechos fundamentales están protegidos y se sigue la normativa europea. Pero ello se asume porque el sistema de IA está implementado en el marco europeo y se ha seguido su normativa. La policía habría de responder sobre si los derechos fundamentales están protegidos y no se ven perjudicados por los sesgos o si los datos personales que emplea el sistema están protegidos en el uso diario de la aplicación. Comentan los creadores que sí que están protegidos, pero habría que analizar cuál es su actuación diaria.

Lo mismo sucedería con la revisión ética. Es la policía quien ha de informar sobre la eticidad en el funcionamiento e implementación de la aplicación. Porque, aunque los creadores estuvieron apoyando la implementación de una revisión ética en las pruebas que se hicieron, así como en su desarrollo, al final los usuarios finales y quienes utilizan diariamente el programa son los que tienen la información clave.

Se establece como posible la realización del monitoreo de los anteriores puntos, pero no podemos afirmarlo con rotundidad porque es la policía quien habría de confirmar si se realiza dicho monitoreo. Se sabe que en la fase de desarrollo y en las etapas previas de funcionamiento e implementación sí era posible, porque se ha proporcionado información sobre las pruebas del algoritmo. Sin embargo, sobre su funcionamiento diario no hay información disponible porque pertenece a la inteligencia interna de la policía.

CAPITULO VII. CONCLUSIONES

La Inteligencia Artificial estará cada vez más integrada en la vida diaria de los ciudadanos, empresas y Administraciones Públicas, el caso de VeriPol, es prueba de ello. Con la consolidación del empleo de estas herramientas en las actividades cotidianas de las organizaciones públicas, el cambio en sus formas de trabajo y dinámicas de relación con los ciudadanos será cada vez más patente, hasta que las transforme, una vez plenamente consolidada la IA.

Por ello se ha de tener presente el marco de conceptualización general que se establece para la IA. No se ha de descartar que pueda haber cambios en su conceptualización en los próximos años, por lo que se ha de estar pendiente a la actualización de las diversas normativas que regulan a estas tecnologías. La IA es una tecnología capaz de analizar una gran cantidad de datos que puede facilitar la actuación de las AA.PP. La relación que ha de darse entre IA y AA.PP. ha de estudiarse y analizarse, este campo se encuentra en sus primeras fases de estudio, pero será de gran relevancia en los próximos años.

Poco a poco se avanza en la electronificación de la Administración para pasar a la smartificación de la Administración. La Administración 4.0. Esta nueva A.P. no solo utiliza los beneficios de lo telemático, sino que además busca facilitar el análisis de datos necesario a gran escala para examinar de una forma más amplia las necesidades reales de los ciudadanos y ofrecer así, unos servicios públicos que se ajusten a sus necesidades. Así como hacer más eficiente el empleo de sus recursos en tareas que los requieren con mayor urgencia. Ello hace que el modelo de Administración cambie por completo y sea necesario un enfoque más holístico capaz de tener en cuenta todos los elementos que inciden en la institución.

La relación que se establece entre la IA y las AA.PP. puede ser muy beneficiosa, pero también se han de considerar los riesgos que acarrea la introducción de estas herramientas en las Administraciones. Uno de los aspectos fundamentales es el relacionado con la gobernanza del dato. Para que la IA pueda desarrollar todas sus preciadas funcionalidades ha de tener una base de datos muy amplia, que en muchos casos incluye datos personales o información de alta sensibilidad que ha de ser tratada con estrictas normas de seguridad. Las AA.PP. habrán de establecer una buena gestión del dato que asiente cómo ha de ser el tratamiento adecuado de la información: cómo se obtiene; dónde se almacena; quién se encarga de su obtención y almacenamiento; si hay una gestión pública, privada o mixta de los datos.

Si se quiere examinar los riesgos que puede presentar la IA para los ciudadanos en la prestación de servicios se han de tener presentes los conceptos básicos de la IA, así como la gestión de los datos y el modelo de administración que se sigue para su adecuada gobernanza. La UE establece siete puntos clave para analizar los posibles riesgos que puede generar el empleo de la IA, los cuales se han seguido para analizar la aplicación de VeriPol en la Policía Nacional española, encargada de detectar denuncias falsas.

Veripol, tal y como ha sido diseñada superaría satisfactoriamente el análisis de las diferentes categorías comentadas. Sin embargo, queda por obtener información sobre el día a día de la aplicación, información que se obtendría desde fuentes policiales. Solo se ha podido acceder a la información disponible en internet, y aunque se comenta que es

una aplicación cuya información es pública, no toda la información sobre la aplicación se encuentra disponible, seguramente por motivos de seguridad. Por ello, para futuras investigaciones se recomienda conseguir realizar entrevistas a policías a cargo de VeriPol o fuentes cercanas a la policía que pudieran responder a preguntas sobre la aplicación.

A raíz del análisis realizado de la aplicación se muestra la necesidad de realizar estudios sobre este tipo de servicios que emplean la IA. Si bien en las primeras fases, tanto en su diseño como primeros testeos, se puede contrastar y afirmar que respondía de forma satisfactoria a las categorías de análisis establecidas en base a lo establecido por la UE. En un principio se supone que la Policía Nacional sigue el mismo protocolo y las mismas acciones que en su momento se siguieron en las pruebas de VeriPol. Pero no se puede afirmar con evidencias sólidas que la policía siga actualmente dichos procedimientos de monitorización.

Este hecho pone de manifiesto que no toda la información sobre VeriPol es pública y accesible. Se puede acceder a los datos básicos de la aplicación e incluso su algoritmo. Pero a la hora de preguntarse cuáles son los posibles riesgos que esta aplicación con IA puede generar para los ciudadanos se requiere una información más completa sobre dicha herramienta y el entorno que la rodea. No solo se necesitan los datos o características básicas de la aplicación, sino saber cómo esta funciona en su día a día para saber si se cumple con los procesos de monitorización pertinentes para solventar los riesgos que VeriPol puede suponer para la ciudadanía.

Cierto es que esta herramienta no tiene un contacto directo con los ciudadanos por lo que, en sí, la propia aplicación, no supone una "amenaza" para los ciudadanos, pero el modo en el que sean tratados los datos que emplea VeriPol, así como las decisiones que tome el policía en base a sus análisis, afectan de forma directa a la población. Como comentaba Liberatore, el texto de la denuncia solía contener un mayor sesgo que el que podía generar VeriPol, por ello hay que prestar especial atención a cómo el policía transcribe la denuncia y cómo el policía encargado de VeriPol introduce el texto en la aplicación. Porque, aunque la aplicación pueda generar un determinado sesgo, la información introducida no ha de estar previamente sesgada. Esto puede suponer un problema porque si la información es previamente tratada, a lo mejor el resultado que ofrece VeriPol no es del todo correcto.

A los problemas que pueden generarse con la transcripción inexacta de las declaraciones y los sesgos, se añade el problema del *mathwashing*. En el momento en el que los policías confíen sin dudar en los resultados de VeriPol y no realicen una supervisión adecuada de la conclusión del análisis, se puede estar incurriendo en *mathwhasing* y la decisión final de los policías no sería autónoma, sino que estaría totalmente supeditada a la resolución de la aplicación. Situación que no es deseable.

Lo beneficioso de la implementación de VeriPol es que permite analizar y cribar mucho más rápido las denuncias interpuestas, para así dedicar una mayor cantidad de recursos a casos que realmente lo necesiten. Facilita la toma de decisiones y permite una mejor y más rápida actuación. La aplicación de VeriPol permite una mejor redistribución de los recursos asignados, sobre todo el económico como se ha comentado, pero esto no puede ser posible si no se realiza una formación completa en las comisarías de policía para que los policías sepan cómo utilizar la herramienta.

Como se ha comentado la administración ha de entenderse de forma holística y atender a todo aquello que interviene en el proceso de prestación de servicios, no solo el servicio. Si solo se analiza la prestación del servicio, se estarían obviando otros problemas

que también son de suma relevancia y que han de ser tenidos en cuenta para la buena calidad y eficacia en la prestación del servicio.

Para próximas investigaciones se recomienda seguir las categorías de análisis planteadas, puesto que se basan en lo establecido por la Unión Europea. Las categorías parecen ser satisfactorias, aunque se recomendaría para responderlas un análisis cualitativo y cuantitativo para tener una visión holística de la integración de la aplicación en la organización. Cualitativo, porque es necesario investigar la información que está disponible de forma pública sobre la aplicación, tanto a nivel general como especializado y realizar entrevistas a personas cercanas a VeriPol (los desarrolladores o miembros de la policía). Y cuantitativo, puesto que también es necesario que se realice un conteo, al igual que en VioGén, del número de denuncias en las que ha intervenido VeriPol; así como el porcentaje de uso de la aplicación, ya que, aunque esté disponible en las comisarías, puede que no se utilice en la práctica diaria de la comisaría.

Una vez realizado el análisis de los posibles riesgos que pueden presentarse para los ciudadanos por el empleo de la IA, se ha de afirmar que cómo resultado de la investigación se han obtenido más preguntas que respuestas. Cierto es que es una aplicación que lleva relativamente poco tiempo y se habrá de observar su evolución. Pero se han destacado en las conclusiones una serie de puntos relevantes que habrían de ser tenidos en cuenta para una mejora en la protección de datos de los ciudadanos y de derecho a la información sobre el tratamiento que reciben sus datos personales. Seguramente no se comparte información sobre la aplicación para preservar su seguridad, pero han de encontrarse vías de que se pueda compartir información sin amenazar la seguridad de VeriPol.

CAPÍTULO VI. BIBLIOGRAFÍA

Abeliuk, A., & Gutiérrez, C. (2021). Historia de la inteligencia artificial. *Revista Bits de Ciencia, 21*, 14-21.

AI HLEG. (2019). *Directrices Éticas para una IA Fiable.* Bruselas: Comisión Europea.

AI HLEG. (2020). *The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self assessment.* European Commission, Directorate-General for Communications Networks, Content and Technology. Brussels: Publications Office.

Algorithm Watch. (27 de octubre de 2020). *Spanish police plan to extend use of its lie-detector while efficacy is unclear*. Recuperado el 15 de julio de 2023, de AlgorithmWatch: https://algorithmwatch.org/en/spain-police-veripol/

Allo, P. (2018). Mathematical values and the epistemology of data practices. En E. Bayamlioglu, I. Baraliuc, L. Janssens, & M. Hildebrandt, *Being Profiled: Cogitas Ergo Sum. 10 Years of 'Profiling the European Citizen'* (págs. 20-23). Amsterdam: Amsterdam University Press.

Almunia, M., & Rey-Biel, P. (2020). *Por un cambio de cultura en la gestión de datos en España: Una propuesta de reforma.* ESADE. Recuperado el 18 de julio de 2023, de https://itemsweb.esade.edu/research/EsadeEcPol_Insight17_Cambiocultura.pdf

Álvarez, R. (13 de abril de 2019). La inteligencia artificial de la Policía que desenmascara denuncias falsas. *La Vanguardia*. Recuperado el 15 de julio de 2023, de https://www.lavanguardia.com/tecnologia/20190414/461583468024/veripol-policia-nacional-inteligencia-artificial-algoritmo-denuncias-falsas.html

Arenilla, M. (2021). *La Administración digital. Los riesgos de la desintermediación, las escisiones y las centralizaciones.* Madrid: INAP.

Asilomar. (2016). Los 23 principios de la Conferencia Asilomar para un uso beneficioso de la Inteligencia Artificial. Recuperado el 1 de junio de 2023, de https://www.triforminstitute.com/wp-content/uploads/2016/12/ASILOMAR.pdf

Asimov, I. (1984). *Yo, robot.* Barcelona: Edhasa.

Barrera, L. (julio-diciembre de 2012). Fundamentos Históricos y Filosóficos de la Inteligencia Artificial. *Revista de Investigación y Cultura, 1*(1), 87-92.

Berryhill, K., Kok Heang, K., Clogher, R., & McBride, K. (2020). *Hola, Mundo: La Inteligencia Artificial y su uso en el Sector Público. Documentos de Trabajo de la OCDE sobre Gobernanza Pública.* Ciudad de México: OCDE .

BigData_uc3m. (26 de enero de 2022). *Applications of AI and Data Science in Policing: 7 years of collaborations with the Spanish Police [vídeo]*. YouTube. Recuperado el 20 de julio de 2023, de https://www.youtube.com/watch?v=z8uBNNPtUmE

Borrajo, D., Juristo, N., Martínez, V., & Pazos, J. (1997). *Inteligencia Artificial. Métodos y técnicas.* Madrid: Centro de estudios Ramón Areces, S.A.

Campos, M. C. (2019). Inteligencia Artificial e Innovación en la Administración Pública: (in)necesarias regulaciones para la garantía del servicio público. *Revista Vasca de Gestión de Personas y Organizaciones Públicas, 3*, 74-91.

Coeckelbergh, M. (2023). *La filosofía política de la inteligencia artificial. Una introducción.* Madrid: Cátedra.

Comisión Europea. (2020a). *AI Watch Artificial Inteligence in public servicies. Overview of the use and impact of AI in public servicies in the EU.*

Comisión Europea. (2020b). *Libro Blanco sobre la inteligencia artificial- un enfoque europeo orientado a la excelencia y la confianza.* Obtenido de commission-white-paper-artificial-intelligence-feb2020_es.pdf (europa.eu)

Corbetta, P. (2010). *Metodología e investigación.* Madrid: McGraw-Hill.

Cortina, A. (7 de mayo de 2019). Ética de la Inteligencia Artificial. *Proyecto de Investigación Científica y Desarrollo Tecnológico*, 379-394.

Corvalán, J. C. (mayo/agosto de 2017). Adminsitración Pública digital e inteligente: transformación en la era de la inteligencia artificial. *Revista de direito Econômico e Socioambiental, 8*(2), 26-56.

Criado, J. I. (2021). Inteligencia Artificial (y Administración Pública). *Eunomía. Revista en Cultura de la Legalidad, 20*, 248-372.

De la Sierra, S. (2020). Inteligencia Artificial y Justicia Administrativa: una aproximación desde la Teoría de Control de la Administración Pública. *Revista General de Derecho Administrativo*(53), 1-19.

Dirección General de Coordinación y Estudios. (2023). *Estadística del Sistema de Seguimiento Integral en los Casos de Violencia de Género (Sistema VioGén).* (M. d.-S. Seguridad, Ed.) Obtenido de INE.

Dirección General de la Policía . (27 de 10 de 2018). *La Policía Nacional pone en funcionamiento la aplicación informática VeriPol para detectar denuncias falsas*. Recuperado el 22 de 08 de 2023, de Policía Nacional: https://www.policia.es/_es/comunicacion_prensa_detalle.php?ID=4433&idiomaActual=es#

Dirección General de la Policía. (2021). *Policía Nacional*. Recuperado el 22 de 08 de 2023, de Sidenpol y Partes de intervención: https://sede.policia.gob.es/portalCiudadano/_es/tramites_ciudadania_atestados.php#

ESSC. (2020). *Item 4 of the agenda. Actions enabling the use of privately held data for official statistics. Work Programme Objective [7] Reaping the benefits of data revolution and moving to trusted smart statistics.* Luxemburgo. Recuperado el 19

de julio de 2023, de https://cros-legacy.ec.europa.eu/system/files/1-essc_2020_43_04_actions_enabling_use_of_privately_held_data_en.pdf.pdf

Etalab. (21 de 05 de 2014). *Open government data: France creates the role of State Chief Data Officer*. Obtenido de Etalb. Politique pblique de la donnée: https://www.etalab.gouv.fr/open-government-data-france-creates-the-role-of-state-chief-data-officer/

Europa Press. (29 de mayo de 2019). La aplicación policial VeriPol, clave para detectar una denuncia falsa en Oviedo. *20minutos*. Recuperado el 15 de julio de 2023, de https://www.20minutos.es/noticia/3655724/0/aplicacion-policial-veripol-clave-para-detectar-denuncia-falsa-oviedo/

Europa Press Asturias. (29 de mayo de 2019). La aplicación policial VeriPol, clave para detectar una denuncia falsa en Oviedo. *Europapress*. Recuperado el 15 de julio de 2023, de https://www.europapress.es/asturias/noticia-aplicacion-policial-veripol-clave-detectar-denuncia-falsa-oviedo-20190529160053.html

EY. (2020). *Inteligencia Artificial en el Sector Público. España. Perspectivas europeas para 2020 y años siguientes.* Madrid: Ernst & Young LLP.

Filgueiras, F. (2021). Inteligencia Artificial e Innovación en la Administración Pública: ambigüedad y elección de sistemas de IA y desafíos de gobernanza digital. *Revista del CLAD Reforma y Democracia*(79).

France 24 Español. (30 de noviembre de 2018). *VeriPol: un algoritmo para detectar las denuncias falsas en España [vídeo]*. YouTube. Recuperado el 15 de julio de 2023, de https://www.youtube.com/watch?v=eAfw4Rg77zA

Gobierno de España. (2020). *Estrategia Nacional de Inteligencia Artificial.* Ministerio de Asuntos Económicos y Transformación Digital, Madrid.

Gobierno de España. (16 de 06 de 2021). *Plan de Recuperación, Transformación y Resilencia. Componenete 16. Estrategia nacional de Inteligencia Artificial.* Presidencia del Gobierno, Madrid.

HLEG BGDS. (2020). *Towards a European strategy on business-to-goverment data sharing for the public interest: Final report prepared by the High-Level Expert Group on Business-to-Government Data Sharing.* Luxemburgo: Publications Office of the European Union.

House of the Commons. (2017). *Robotics and artificial intelligence.* Science and Technology. Londres: Order of the House.

INE. (2023a). *Estadística de condenados: Adultos. Hurtos y Robos.* Recuperado el 30 de julio de 2023, de Instituto Nacional de Estadística: https://www.ine.es/jaxiT3/Datos.htm?tpx=58531#!tabs-grafico

INE. (2023b). *Estadística de condenados: Adultos. Acusación y denuncia falsas y simulación de delitos*. Recuperado el 30 de julio de 2023, de Instituto Nacional de Estadística: https://www.ine.es/jaxiT3/Datos.htm?tpx=58531#!tabs-grafico

La Moncloa. (27 de octubre de 2018). *La Policía Nacional pone en funcionamiento la aplicación informática VeriPol para detectar denuncias falsas*. Recuperado el 22 de 07 de 2023, de La Moncloa: https://www.lamoncloa.gob.es/serviciosdeprensa/notasprensa/interior/Paginas/2018/271018veripol.aspx

Lee, K. F. (2018). *AI superpowers: China, Silicon Valley, and the new world order.* Boston: Houghton Mifflin Harcourt.

Leonoticias. (18 de enero de 2019). Identificada una vecina de León que presentó una denuncia falsa sobre el robo de su móvil. *Leonoticias*. Recuperado el 15 de julio de 2023, de https://www.leonoticias.com/leon/identificada-vecina-leon-20190118134753-nt.html

Ministerio del Interior. (14 de diciembre de 2017). *El Secretario de Estado de Seguridad ha presidido la entrega de los Premios de Investigación de la Fundación Policía Española*. Recuperado el 13 de mayo de 2023, de Ministerio del Interior: https://www.interior.gob.es/opencms/es/detalle/articulo/El-secretario-de-Estado-de-Seguridad-ha-presidido-la-entrega-de-los-Premios-de-Investigacion-de-la-Fundacion-Policia-Espanola/

Ministerio del Interior. (27 de octubre de 2018). *La Policía Nacional pone en funcionamiento la aplicación informática VeriPol para detectar denuncias falsas* . Recuperado el 22 de 07 de 2023, de Ministerio del Interior: https://www.interior.gob.es/opencms/es/detalle/articulo/La-Policia-Nacional-pone-en-funcionamiento-la-aplicacion-informatica-VeriPol-para-detectar-denuncias-falsas/

Misuraca, G., & Van Noordt, C. (2020). *Overview and impact of AI in public services in the EU*. Luxembourg: Publications office of the European Union.

Muñoz, R. (2020). Las TICS en la Administración Pública. La Inteligencia Artificial ante una perspectiva de derechos. *Editoril Astrea*, 1-24.

Norvig, P., & Russell, S. (2010). *Artificial Intelligence. A modern approach.* Upper Saddle River: Pearson.

Noticias de Almería. (06 de junio de 2021). La app Veripol desvela la falsa denuncia de 7.000 euros y acaba detenido el denunciante. *Noticiasde almería.com*. Recuperado el 15 de julio de 2023, de https://www.noticiasdealmeria.com/la-app-veripol-desvela-la-falsa-denuncia-de-7.000-euros-y-acaba-detenido-el-denunciante

Ocaña-Fernández, Y., Valenzuela, L. A., Vera-Flores, M. A., & Rengifo-Lozano, A. (2021). Inteligencia Artificial (IA) aplicada a la gestión pública. *Revista Venezolana de Gerencia, 26*(94), 696-504.

OCDE/Eurostat. (2018). *Oslo Manual: Guidelines for Collecting, Reporting and Using Data on Innovation 4th Edition. The Measurement of Scientific, Technological and Innovation Activities.* Paris/Eurostat. Luxembourg: OECD Publishing.

OHCHR. (10 de mayo de 2022). *La Inteligencia Artificial y los Objetivos de Desarrollo Sostenible*. Recuperado el 18 de junio de 2023, de ONU. Novedades/Artículos. Objetivos de desarrollo sostenible: https://www.ohchr.org/es/stories/2022/05/artificial-intelligence-and-sustainable-development-goals

Oliver, N. (2020). *Inteligencia Artificial, naturalmente, un manual de convivencia entre humanos y máquinas para que la tecnología nos beneficie a todos. Pensamiento para la sociedad digital nº1*. Observatorio Nacional de las Telecomunicaciones.

ONU. (25 de noviembre de 2021). *Primer acuerdo mundial sobre la ética de la inteligencia artificial*. Recuperado el 20 de julio de 2023, de Noticias ONU. Mirada global. Historias Humanas: https://news.un.org/es/story/2021/11/1500522

OSPI. (2017). *Inteligencia Artificial y su aplicación en los Servicios Públicos. Documento de Conclusiones.* Observatorio Público IECISA. El Corte Inglés S.A.

OTRI. (2018). *Veripol, inteligencia artificial a la caza de denuncias falsas.* Madrid: Unidad de Cultura y Divulgación. Universidad Complutense de Madrid.

Parlamento Europeo. (04 de mayo de 2022). *Inteligencia artificial: oportunidades y desafíos*. Recuperado el 07 de junio de 2023, de Noticias. Parlamento Europeo: https://www.europarl.europa.eu/news/es/headlines/society/20200918STO87404/inteligencia-artificial-oportunidades-y-desafios

Parlamento Europeo. (06 de junio de 2023). *Ley de IA de la UE: primera normativa sobre inteligencia artificial*. Recuperado el 17 de junio de 2023, de Noticias. Parlamento Europeo.: https://www.europarl.europa.eu/news/es/headlines/society/20230601STO93804/ley-de-ia-de-la-ue-primera-normativa-sobre-inteligencia-artificial

Parlamento Europeo. (15 de junio de 2023). *Regulación de la inteligencia artificial en la UE: la propuesta del Parlamento*. Obtenido de Noticias. Parlamento Europeo.: https://www.europarl.europa.eu/news/es/headlines/society/20201015STO89417/regulacion-de-la-inteligencia-artificial-en-la-ue-la-propuesta-del-parlamento

Peters, G. B., & Pierre, J. (2005). ¿Gobernanza sin gobierno? Replanteándose la Administración Pública. En A. Cerrillo, *La gobernanza hoy: 10 textos de referencia* (págs. 123-144). Madrid: Instituto Nacional de Administración Pública .

Pita, E. (30 de agosto de 2020). La comisaría de Vigo lleva dos años sin usar la «máquina de la verdad». *La Voz de Galicia*. Recuperado el 15 de julio de 2023, de https://www.lavozdegalicia.es/noticia/vigo/vigo/2020/08/30/comisaria-vigo-lleva-dos-anos-usar-maquina-verdad/0003_202008V30C1991.htm

Portillo, E. (11 de octubre de 2018). *La policía española implanta el primer sistema capaz de detectar mentiras falsas*. Recuperado el 02 de julio de 2023, de Red Estratégica en Matemáticas: https://institucionales.us.es/remimus/la-policia-espanola-implanta-el-primer-sistema-capaz-de-detectar-mentiras-falsas/

Quijano, L., Liberatore, F., Camacho, M., & Camacho, J. (1 de junio de 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*, 155-168.

R.G. (08 de julio de 2019). Esclarecidas más de 40 denuncias falsas en el primer semestre del año en Granada, Baza y Motril. *Granada Hoy*. Recuperado el julio de 15 de 2023, de https://www.granadahoy.com/provincia/Esclarecidas-denuncias-falsas-primer-semestre_0_1371163327.html

Ramió, C. (2018). Inteligencia Artificial, robótica y modelos de Administración pública. *Revista CLAD Reforma y Democracia*(72), 5-42.

Ramió, C. (2019). *Inteligencia Artificial y Adminsitración Pública. Robots y humanos compartiendo el servicio público.* Madrid: Catarata.

Redacción Salmanca24horas. (07 de junio de 2019). Descubierta una mujer que puso una denuncia falsa sobre el robo de su bolso gracias a la aplicación 'VeriPol'. *Salmanca24horas.com*. Recuperado el 15 de julio de 2023, de https://www.salamanca24horas.com/sucesos/descubierta-mujer-puso-denuncia-falsa-sobre-robo-bolso-gracias-aplicacion-veripol_1446841_102.html

RTVE. (09 de noviembre de 2018). *VERIPOL, el programa policial que detecta denuncias falsas [vídeo]*. RTVEPlay. Recuperado el 15 de julio de 2023, de https://www.rtve.es/play/videos/programa/td1-veripol-021118/4822077/

Salvador, M. (julio de 2021). Inteligencia artificial y gobernanza de datos en las administraciones públicas: reflexiones y evidencias para su desarrollo. *Nueva Época*(26), 20-32.

Salvador, M., & Ramió, C. (2020). Capacidades analíticas y gobernanza de datos en la administración pública como paso previo a la introducción de la inteligencia artificial. *Revista CLAD Reforma y Democracia*(77), 5-36.

Sindicato CPPM. (28 de octubre de 2018). *VeriPol: así es el nuevo 'programa' de la Policía Nacional capaz de detectar denuncias falsas. Telenoticias Madrid. [vídeo]*. YouTube. Obtenido de https://www.youtube.com/watch?v=WMvTfKYXlHY

Telefónica, F. (2021). *Sociedad digital en España. El año en que todo cambió 2020-2021*. Barcelona: Penguin Rnadom House Grupo Editorial .

Villanueva, C. (2020). El estado regulador desde la perspectiva del derecho a la buena administración y la ética pública. En A. Fernández, F. Morandini, & M. Gonzalez, *La Buena Administración y la Ética Pública en el Derecho Administrativo* (págs. 199-244). Grupo editorial Ibañez.

Weber, M. (1947). *The Theory Of Social And Economic Organization.* Glencoe: The Falcon Wing´s Press.

Willems, J., Schmid, M., Vanderelest, D., Vogel, D., & Ebinger, F. (2022). AI- dirven public services and the privacy paradox. do citizens really care about their privacy? *Public Management Review*, 1-19.

ANEXOS

ANEXO 1

Principios de Asimov

1. Un robot no debe dañar a un ser humano o, por su inacción, dejar que un ser humano sufra daño.
2. Un robot debe obedecer las órdenes que le son dadas por un ser humano, excepto cuando estas órdenes se oponen a la primera Ley.
3. Un robot debe proteger su propia existencia, hasta donde esta protección no entre en conflicto con la primera o segunda Leyes.

Recuperado de (Asimov, 1984)

ANEXO 2

Los 23 principios de la Conferencia Asilomar para un uso beneficioso de la Inteligencia Artificial.

La inteligencia artificial ya ha proporcionado herramientas beneficiosas que son utilizadas diariamente por personas de todo el mundo. Su continuo desarrollo, guiado por los siguientes principios, ofrecerá increíbles oportunidades para ayudar y empoderar a las personas en las décadas y siglos venideros.

Temas de investigación

1) Objetivo de la investigación: El objetivo de la investigación de la IA debe ser crear no inteligencia no dirigida, sino inteligencia beneficiosa.

2) Financiación de la investigación: Las inversiones en AI deben ir acompañadas de fondos para la investigación que asegure su uso beneficioso, incluyendo cuestiones espinosas en ciencias de la computación, economía, derecho, ética y estudios sociales, tales como:

¿Cómo podemos hacer que los futuros sistemas de IA sean altamente robustos, para que hagan lo que queramos sin que funcionen mal o sean pirateados?

¿Cómo podemos aumentar nuestra prosperidad a través de la automatización mientras mantenemos los recursos y el propósito de las personas?

¿Cómo podemos actualizar nuestros sistemas legales para ser más justos y eficientes, para mantener el ritmo de la IA y para manejar los riesgos asociados con la IA?

¿Con qué conjunto de valores debería alinearse la IA y qué estatus legal y ético debería tener?

3) Vínculo entre ciencia y política: Debe haber un intercambio constructivo y saludable entre los investigadores de la IA y los responsables de la formulación de políticas.

4) Cultura de la investigación: Se debe fomentar una cultura de cooperación, confianza y transparencia entre los investigadores y desarrolladores de la IA.

5) Evasión de razas: Los equipos que desarrollen sistemas de IA deben cooperar activamente para evitar recortes en las normas de seguridad.

Ética y Valores

6) Seguridad: Los sistemas de IA deben ser seguros durante toda su vida útil, y verificables siempre que sea posible y factible.

7) Fallo Transparencia: Si un sistema de IA causa daño, debería ser posible determinar por qué.

8) Transparencia Judicial: Toda participación de un sistema autónomo en la toma de decisiones judiciales debe proporcionar una explicación satisfactoria que pueda ser auditada por una autoridad humana competente.

9) Responsabilidad: Los diseñadores y constructores de sistemas avanzados de IA son partes interesadas en las implicaciones morales de su uso, mal uso y acciones, con la responsabilidad y oportunidad de dar forma a esas implicaciones.

10) Alineación de valores: Los sistemas de IA altamente autónomos deben ser diseñados de manera que sus objetivos y comportamientos puedan ser asegurados para alinearse con los valores humanos a lo largo de su operación.

11) Valores Humanos: Los sistemas de IA deben ser diseñados y operados de manera que sean compatibles con los ideales de dignidad humana, derechos, libertades y diversidad cultural.

12) Privacidad personal: Las personas deben tener el derecho de acceder, administrar y controlar los datos que generan, dado el poder de los sistemas de IA para analizar y utilizar esos datos.

13) Libertad y Privacidad: La aplicación de la IA a los datos personales no debe restringir irrazonablemente la libertad real o percibida de las personas.

14) Beneficio Compartido: Las tecnologías de IA deben beneficiar y empoderar a tantas personas como sea posible.

15) Prosperidad compartida: La prosperidad económica creada por AI debe compartirse ampliamente, para beneficiar a toda la humanidad.

16) Control Humano: Los humanos deben elegir cómo y si delegar decisiones a los sistemas de IA, para lograr los objetivos elegidos por el ser humano.

17) No subversión: El poder que confiere el control de sistemas de inteligencia artificial muy avanzados debe respetar y mejorar, en lugar de subvertir, los procesos sociales y cívicos de los que depende la salud de la sociedad.

18) Carrera armamentista AI: Debe evitarse una carrera armamentista con armas autónomas letales.

Temas a largo plazo

19) Capacidad Precaución: Al no haber consenso, debemos evitar las suposiciones firmes con respecto a los límites máximos de las futuras capacidades de IA.

20) Importancia: La IA avanzada podría representar un cambio profundo en la historia de la vida en la Tierra, y debería planificarse y gestionarse con el cuidado y los recursos adecuados.

21) Riesgos: Los riesgos que plantean los sistemas de IA, especialmente los riesgos catastróficos o existenciales, deben estar sujetos a esfuerzos de planificación y mitigación acordes con su impacto esperado.

22) Auto-mejora Recursiva: Los sistemas de inseminación artificial diseñados para mejorar o replicarse recursivamente de manera que puedan conducir a un rápido aumento de la calidad o la cantidad deben estar sujetos a estrictas medidas de seguridad y control.

23) Bien Común: La Superinteligencia sólo debe desarrollarse al servicio de ideales éticos ampliamente compartidos, y en beneficio de toda la humanidad y no de un solo Estado u organización.

Recuperado de (Asilomar, 2016)

ANEXO 3

El lun, 24 jul 2023, 15:58, Elena Del Castillo Gil <e.delcastillo.2016@alumnos.urjc.es> escribió:

Buenas tardes,

Soy Elena del Castillo Gil, alumna del grado en Ciencias Políticas de la URJC y estoy realizando mi TFG sobre los riesgos que supone para el ciudadano la implementación de la IA en los servicios públicos. En concreto, he elegido el caso de VeriPol.

Usted fue una de las personas que participó en la elaboración del proyecto y quería preguntarle si sería posible realizarle una entrevista – online o presencial- para la investigación que estoy llevando acabo en mi TFG.

Muchas gracias por su tiempo,

Espero su respuesta,

Un saludo,

Elena

Re: TFG SOBRE VERIPOL URJC

Lara Quijano Sanchez <lara.quijano@uam.es>

Mar 25/07/2023 14:32

Para:

Hola,

Me alegra estes estudiando IA desde tu área. Es muy bueno tener carácter multidisciplinario. Creo la IA puede ayudar muchísimo a la ciudadanía y tiene un enorme potencial y ventajas. Ahora mismo estoy de permiso por maternidad.

Pero te paso esta charla que di sobre veripol explicando cada detalle y también las ventajas que la IA puede tener.

https://youtu.be/z8uBNNPtUmE

Entiendo que el tema científico de tu tfg es una hipótesis de <si existe riesgo o no > donde lo evaluaras. Seguro el video te sirve. Para poder evaluarlo seguro necesitas también cursos para entender la IA y los algoritmos por dentro y como funcionan te recomiendo este curso aunque hay mas.

https://www.fundacion.uc3m.es/formacion/curso-en-inteligencia-artificial-ciencia-de-datos/

Mucha suerte.

Saludos,

ANEXO 4