Digital Future Society

El uso de algoritmos en el sector público en España: cuatro estudios de caso sobre ADMS

Un programa de

GOBIERNO DE ESPAÑA
VICEPRESIDENCIA PRIMERA DE GOBIERNO
MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL
SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

red.es

MOBILE WORLD CAPITAL BARCELONA

Sobre Digital Future Society

Digital Future Society es una iniciativa transnacional sin ánimo de lucro que conecta a responsables políticos, organizaciones cívicas, expertos académicos y empresarios para explorar, experimentar y explicar cómo se pueden diseñar, usar y gobernar las tecnologías a fin de crear las condiciones adecuadas para una sociedad más inclusiva y equitativa.

Nuestro objetivo es ayudar a los responsables políticos a identificar, comprender y priorizar los desafíos y las oportunidades fundamentales, ahora y en los próximos diez años, en relación con temas clave que incluyen la innovación pública, la confianza digital y el crecimiento equitativo.

Para más información, visite digitalfuturesociety.com

Un programa de

GOBIERNO DE ESPAÑA
VICEPRESIDENCIA PRIMERA DE GOBIERNO
MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL
SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

red.es

MOBILE WORLD CAPITAL BARCELONA

Publicado
Febrero del 2023

Digital Future Society

Contenidos

Introducción

Hay quienes presentan la inteligencia artificial (IA) como una solución mágica a problemas sociales complejos, pero también quienes la consideran un instrumento para justificar el aumento de la vigilancia y la datificación de nuestras sociedades actuales. En este contexto, **los Gobiernos se están apresurando a implantar tecnologías basadas en datos, entre ellas los sistemas automatizados de toma de decisiones (ADMS)**, que ayuden al sector público en tareas como la optimización de procesos internos, la previsión de riesgos y la asignación de recursos.

El uso de la IA por parte de las Administraciones públicas es complejo y comporta multitud de aspectos y matices. A medida que se generaliza el uso de herramientas basadas en datos para ayudar a la gobernanza, **es fundamental que tanto las Administraciones públicas como los ciudadanos comprendan las oportunidades, los retos y los riesgos que conlleva su uso**.

Organizaciones de la sociedad civil, el ámbito académico y los organismos reguladores han advertido sobre las cuestiones éticas que implica el desarrollo y el uso de la IA. Tras la pandemia de la COVID-19, es aún más importante abordar la digitalización del sector público con un enfoque integral, dado que los países que se están recuperando pueden verse cegados por la necesidad urgente de adoptar soluciones digitales.

La promesa de un sector público impulsado por la tecnología, más ágil y eficiente, ha sido una narrativa habitual en la modernización de las Administraciones de todo el mundo.

Siguiendo dicha narrativa, en su estrategia digital para el 2025, el Gobierno español destaca el sector público como una de las principales áreas en las que acelerará la digitalización. Según esa estrategia, las tecnologías emergentes, como la inteligencia artificial (IA), son clave para mejorar la eficiencia y la calidad de sus servicios y velar por que los ciudadanos con competencias digitales puedan interactuar de forma segura con la Administración pública.

En ese sentido, España ha destinado 4.000 millones de euros de los fondos Next Generation EU al Plan Nacional de Recuperación y Resiliencia para hacer reformas e invertir en la creación de "una Administración para el siglo XXI" (Ministerio de Asuntos Económicos y Transición Digital s. f.; Mileusnic 2022).

Por otra parte, la IA, los algoritmos y la automatización se están incorporando cada vez más al vocabulario cotidiano. A medida que aumenta el uso de la IA, crece la necesidad de comprender de qué manera, exactamente, pueden tener un impacto directo en la vida de las personas la digitalización de los servicios públicos y el uso de la automatización. Una mera equivocación al marcar una casilla en una solicitud online puede hacer que a todo un grupo de personas se le deniegue una ayuda a la que tiene derecho; un error de cálculo en un control policial puede ser un factor de vida o muerte; una alerta pasada por alto puede dar

lugar a un mal uso del dinero de los contribuyentes. **Cuanto más se delegan las decisiones en algoritmos, más necesitamos comprender cómo se toman esas decisiones y qué información se utiliza para tomarlas.**

Este informe se basa en los trabajos previos de Digital Future Society sobre los sistemas automatizados de toma de decisiones, y continúa profundizando en la comprensión de su impacto social.

También continúa el esfuerzo de Digital Future Society por entender el impacto social que se deriva del uso de los ADMS. Por lo tanto, este informe no trata los aspectos técnicos de los ADMS, sino que se centra en los ADMS como sistemas sociotécnicos, es decir, sistemas que están impregnados de los valores, perspectivas y sesgos de las personas que los crean y los utilizan.

Además, pretende servir de puente entre los conocimientos de diferentes expertos del mundo académico, el sector público y las organizaciones de la sociedad civil, para ilustrar la complejidad que supone cada herramienta.

Por último, en nuestro trabajo y en este informe, definimos los ADMS como sistemas que refuerzan o reemplazan un proceso de toma de decisiones que de otro modo sería realizado por humanos. En la literatura y también en este informe, los términos ADMS e IA se usan indistintamente, al considerarse los ADMS una subcategoría de la IA.

El presente informe explora cuatro estudios de caso de ADMS utilizados por el sector público español que han llamado la atención de los medios de comunicación. Son los siguientes:

BOSCO

Una herramienta que determina si se tiene derecho a una subvención nacional de la factura eléctrica.

RisCanvi

Un sistema que calcula el riesgo de reincidencia de los reclusos en Cataluña.

VioGén

Un mecanismo de la Policía Nacional que predice el riesgo de volver a ejercer violencia de género.

SALER

Un sistema que prevé casos potenciales de corrupción en Valencia.

Sobre este informe

En el 2020, Digital Future Society Think Tank exploró el uso de los sistemas automatizados de toma de decisiones (ADMS) en el sector público a través de varios informes sobre el estado de bienestar digital y sus implicaciones para la igualdad de género, así como de un libro blanco sobre los riesgos y las oportunidades de la IA en el sector público (Digital Future Society 2020, 2021).

En la misma línea, **el *think tank* produjo un pódcast en castellano, titulado *Algoritmos y Gobiernos***, en el que se examinaron cuatro herramientas utilizadas en las Administraciones públicas españolas.[1] El objetivo del pódcast era concienciar sobre la adopción de ADMS por parte del sector público e ilustrar la complejidad que hay detrás de su diseño, su gobernanza y su uso. El pódcast incluye entrevistas con investigadores, funcionarios públicos, abogados, miembros de organizaciones de la sociedad civil y ciudadanos.

Este informe se basa en el contenido del pódcast, y en los estudios de caso se emplean extractos de las entrevistas grabadas en él, junto con el resultado de otras investigaciones y de la consulta de documentación. Asimismo, el informe —publicado en inglés y en español— trata de salvar la brecha lingüística y de conocimiento, dado que aún hay pocos estudios sobre los ADMS y sus repercusiones en España y otros países. Este trabajo arroja luz sobre el contexto específico de España, recogiendo ideas de los diferentes actores implicados en la implementación de los ADMS, con especial atención a las aplicaciones que afectan directamente al bienestar de la ciudadanía.

También pretende aunar conocimientos entre las comunidades de habla hispana e inglesa. Considerando la escasez de publicaciones que hablan de estos sistemas, es necesario poner en común lo que ocurre en España, aprender de ello y contribuir a ampliar la información generada sobre este tema.

¿Por qué ahora?

La pandemia de la COVID-19 no solo ha acelerado la digitalización en todos los aspectos de la sociedad: también ha revelado la envergadura de la brecha digital. Se está animando a los Gobiernos a que refuercen sus estrategias digitales y mejoren el uso de las tecnologías digitales. Esto se refleja en los fondos Next Generation EU, ideados para ayudar a los Estados miembros de la Unión Europea en su recuperación, para que sean "digitales, más ecológicos y más resilientes" de cara a los retos venideros (Cedefop 2021).

Al mismo tiempo, Europa está sentando un precedente normativo con su propuesta de Ley de IA, para velar por una IA fiable y centrada en las personas. En España, mientras la ley se negocia en el Parlamento, está previsto iniciar un espacio piloto para poner a prueba los requisitos regulatorios de la IA de alto riesgo. En paralelo, se está creando también un organismo de vigilancia: la Agencia Española de Supervisión de la Inteligencia Artificial.

[1] Para obtener más información, visite https://digitalfuturesociety.com/es/algoritmos-gobiernos-podcast/

Estas medidas son muy oportunas, teniendo en cuenta que la Ley de IA forma parte de una tendencia más amplia en las regulaciones. En todo el mundo, las reacciones a dicha Ley han sido diversas. Mientras que algunos aplauden los esfuerzos de la UE, otros consideran que no es suficiente para proteger derechos fundamentales. Por ejemplo, 119 organizaciones han criticado el enfoque de la Ley de IA, basado en los riesgos, afirmando que no aborda adecuadamente los problemas de derechos humanos que surgen con los sistemas de IA (European Digital Rights 2021).

Ahora que los grandes actores están avanzando y dando forma a la IA, es fundamental que los ciudadanos comprendan diferentes casos de uso específicos de la IA y aprendan de ellos. Es un momento crucial en el que las distintas partes interesadas pueden asegurarse de que las tecnologías de IA tengan en cuenta los derechos humanos (Ibid.).

Público

Por último, este informe está dirigido al público general, especialmente a aquellos que quieran informarse sobre cómo se están utilizando las herramientas algorítmicas en el sector público. El informe también es una fuente útil para responsables políticos, miembros de organizaciones de la sociedad civil, activistas y cualquier persona que estudie la cuestión y tenga interés en comprender estas herramientas, en términos generales, desde una perspectiva social.

Estudios de caso

Este informe examina cuatro sistemas automatizados de toma de decisiones que se utilizan en diferentes Administraciones públicas de toda España. **No son representativos de los diversos tipos de usos** que tienen actualmente estos sistemas. Los sistemas de IA se pueden aplicar en una amplia variedad de campos, como la sanidad, las finanzas, la educación, etc. Los casos elegidos para el informe **pertenecen a los ámbitos de los servicios sociales, la policía y la Administración en general.**

Se han seleccionado estos casos porque todos ellos **han atraído la atención de los medios de comunicación** por diferentes motivos, algunos de ellos negativos: por ejemplo, por emplear prácticas cuestionables en materia de protección de datos, por falta de transparencia o por su potencial de exacerbar las desigualdades existentes. Pero, también, por su posible capacidad de promover la eficiencia y la objetividad. Son los siguientes:

Caso nº 1

BOSCO

Determina si se tiene derecho a una subvención nacional de la factura eléctrica.

Caso nº 2

RisCanvi

Calcula el riesgo de reincidencia de los reclusos en Cataluña.

Caso nº 3

VioGén

Predice el riesgo de volver a ejercer violencia de género para la Policía Nacional.

Caso nº 4

SALER

Prevé casos potenciales de corrupción en Valencia.

Este informe **pretende ofrecer una perspectiva global de cada estudio de caso**, a partir de las percepciones más extendidas entre investigadores, funcionarios y usuarios. Su objetivo no es ofrecer información exhaustiva acerca del funcionamiento de cada sistema, sino arrojar luz sobre las grandes cuestiones que rodean a estos casos.

Cada estudio de caso puede leerse como un artículo independiente y está estructurado para ofrecer una explicación a grandes rasgos del funcionamiento del sistema, seguida de una discusión y un análisis. Algunos usos de la IA son más complejos que otros, como las herramientas de evaluación de riesgos; debido al gran volumen de publicaciones sobre este tema y a la complejidad de estos mecanismos, los casos de estudio correspondientes a dichas herramientas, RisCanvi y Viogén, son un poco más largos que los de los otros dos sistemas, BOSCO y SALER. Al final del informe, en la sección “De cara al futuro”, se resumen las cuatro cuestiones transversales observadas en los cuatro estudios de caso.

Caso nº 1: **BOSCO**

Determinación de si se tiene derecho a una subvención de la factura eléctrica

El contexto

En respuesta a la crisis económica, el Gobierno español creó en el año 2009 el bono social, una subvención de la factura eléctrica para los hogares de rentas bajas. Unos años más tarde, al revisar dicha prestación, el Gobierno impuso unas condiciones más estrictas para poder optar a ella. Paralelamente, implementó el uso de un software llamado BOSCO para evaluar a los solicitantes y decidir si se ajustaban a los nuevos criterios establecidos (Kayser-Bril 2019).

¿Qué es BOSCO?

BOSCO es un programa informático creado por la Administración pública en el 2017 para evaluar si los usuarios tienen derecho a recibir el bono social de la factura eléctrica. Las compañías eléctricas, no el Gobierno, administran el programa y se comunican directamente con los solicitantes.

¿Quién puede optar a la ayuda?

El factor decisivo que determina si alguien tiene derecho a la subvención son sus ingresos. Este es el factor básico, pero existen otros factores "amplificadores" que pueden condicionar la cuantía de la subvención, como el hecho de que la persona beneficiaria haya sido o sea víctima de violencia de género o que padezca una discapacidad. Y tienen derecho a la subvención todas las familias numerosas con tres hijos o más, independientemente de sus ingresos, además de algunos pensionistas, como los que perciben pensiones de jubilación o discapacidad, pero no los que reciben una pensión de viudedad.

La cuestión a debate

En el 2018 se rechazaron las solicitudes de más de medio millón de familias. Esto llamó la atención de Civio, una organización sin ánimo de lucro cuya principal misión es asegurarse de que la Administración pública actúe como debe. Cuando supieron del gran volumen de solicitudes rechazadas, asumieron la tarea de ayudar a los solicitantes en el proceso de apelación.

Civio creó un asistente para ayudarlos a determinar si tenían derecho a la subvención. Con ese fin, partieron del decreto-ley en el que se basaba el software BOSCO y tradujeron los requisitos legales a código informático. Al crear esa herramienta, se dieron cuenta de lo complicado que era traducir la ley a código, porque la propia ley tiene ciertas ambigüedades.

Una vez que Civio implementó su propia herramienta, muchos solicitantes descubrieron que cumplían los requisitos según la interpretación de esta organización, pero se les denegaba la ayuda mediante la herramienta oficial. Las alarmas saltaron cuando Civio recibió numerosas llamadas sobre esta contradicción. Y cuando, por fin, pudo acceder a parte de la información técnica, descubrió que había errores que debían corregirse para que los beneficiarios pudieran recibir la ayuda a la que tenían derecho.

Para crear este asistente, Civio solicitó al Gobierno la documentación utilizada. La solicitud pasó por los ministerios implicados y por el Consejo de Transparencia y Buen Gobierno (CTBG). El CTBG proporcionó la información técnica y los casos de uso, pero no facilitó el código fuente. Sin el código fuente, Civio no pudo determinar si la herramienta era defectuosa desde la base.

Los directores de la organización, Eva Belmonte y David Cabo, solicitaron su código fuente para averiguar a qué se debían los errores, pero la petición fue denegada por motivos de derechos de propiedad intelectual. Según el abogado de Civio, Javier de la Cueva, la interpretación actual de la ley permite a la Administración pública desarrollar algoritmos opacos, lo que ha llevado a Civio a recurrir la denegación ante los tribunales (Civio 2019).

Según explica Belmonte, "nos dieron lo que se supone que hacía la funcionalidad de la aplicación, y ahí descubrimos que algunos de los errores que nosotros habíamos detectado al hablar con la gente que pedía el bono social eran reales, como el hecho de que a las viudas se les decía que no tenían derecho a bono social aunque cumplieran los requisitos". Técnicamente, los beneficiarios de una pensión de viudedad no tienen derecho a esta subvención, pero si la solicitan por el criterio de ingresos, sí pueden cumplir con el umbral de renta. "No sabemos si hay más errores ahí dentro, porque no tenemos acceso al código", explica (Digital Future Society 2022c 00:11:30).

El caso dio un giro en junio del 2022, cuando el Consejo de Transparencia cambió de postura y accedió a revelar el código fuente. Actualmente, Civio está esperando la sentencia definitiva sobre el caso.

Para dicha organización, este caso es importante a largo plazo, ya que sentará un precedente en el que se basará la Administración española al abordar futuros casos relacionados con la transparencia de los ADMS. BOSCO es relativamente sencillo en comparación con otros sistemas que se prevé usar más en el futuro, de mayor complejidad, como los que se basan en el aprendizaje automático, por lo que a Civio le preocupa mucho esta cuestión. A medida que crece el uso de ADMS en el sector público, las preocupaciones de Civio tienen que ver con cómo puede anticiparse la sociedad civil a su uso y defender los derechos de la ciudadanía si se produce alguna infracción en el futuro.

BOSCO

El proceso de solicitud

Cuando una persona se entera de que puede optar a la subvención, puede solicitarla a través de las diversas compañías eléctricas que figuran como distribuidoras en el sitio web del Gobierno español. Los solicitantes pueden pedir la subvención por teléfono, correo electrónico o fax. El proceso consiste en rellenar un formulario y facilitar lo siguiente: los documentos de identidad de los miembros de la unidad familiar, el contrato de electricidad y una prueba sobre el lugar de residencia (si procede, también se debe incluir documentación que demuestre la discapacidad o vulnerabilidad).

Una vez enviados estos documentos, la persona recibe una respuesta automática, y, en el plazo de 15 días, debería recibir por correo electrónico o postal una respuesta definitiva sobre si se le concede la subvención.

La experiencia de los usuarios: Mercedes y su asistente social, Nerea

Mercedes trabaja y es madre soltera con dos hijos a su cargo, uno de los cuales tiene una discapacidad. Gana 1.000 euros al mes y paga 550 de alquiler, además de las cuotas de un préstamo que pidió para pagar la fianza de su piso. “Si encima tengo que pagar el recibo de la luz y el recibo del agua, me quedo sin comer”, afirma (Digital Future Society 2022c 00:00:35).

Mercedes cumple los requisitos para recibir la subvención de la factura eléctrica, pero lo supo gracias a Insercoop, una organización sin ánimo de lucro de L’Hospitalet de Llobregat, ciudad del área metropolitana de Barcelona. Insercoop ayuda a personas vulnerables a encontrar un empleo remunerado y presta asistencia a sus usuarios para realizar tareas burocráticas, especialmente aquellas que requieren competencias digitales, de modo que puedan recibir prestaciones. Esta organización desempeña un papel fundamental a la hora de informar a su público sobre los diferentes tipos de ayudas, ya que muchos no saben que podrían recibir asistencia.

La asistente social de Mercedes, Nerea, la ayudó a rellenar su solicitud y a enviarla a la compañía eléctrica. Pero, tras los 15 días establecidos, no recibieron la respuesta habitual. Nerea llamó y le informaron de que faltaba documentación en la solicitud y no podían tramitarla. Días después de la llamada, Mercedes recibió una carta por correo postal, informándole de que faltaban documentos. Tuvo que esperar más de tres semanas para recibir una respuesta definitiva.

Mercedes explica que, sin Nerea, se sentiría perdida y abrumada por el proceso de solicitud. Además de que carece de las competencias digitales necesarias para solicitar las prestaciones, no es fácil orientarse ante multitud de servicios, como atestigua Nerea. Las personas con competencias digitales, como enviar correos electrónicos o subir documentos escaneados, tienen ventaja a la hora de solicitar ayudas, ya que la asistencia telefónica es un proceso tedioso y que requiere mucho tiempo, algo que Mercedes no puede permitirse.

BOSCO es un programa informático que usan las compañías eléctricas. Ha sido desarrollado por el Gobierno español para evaluar las solicitudes del bono social

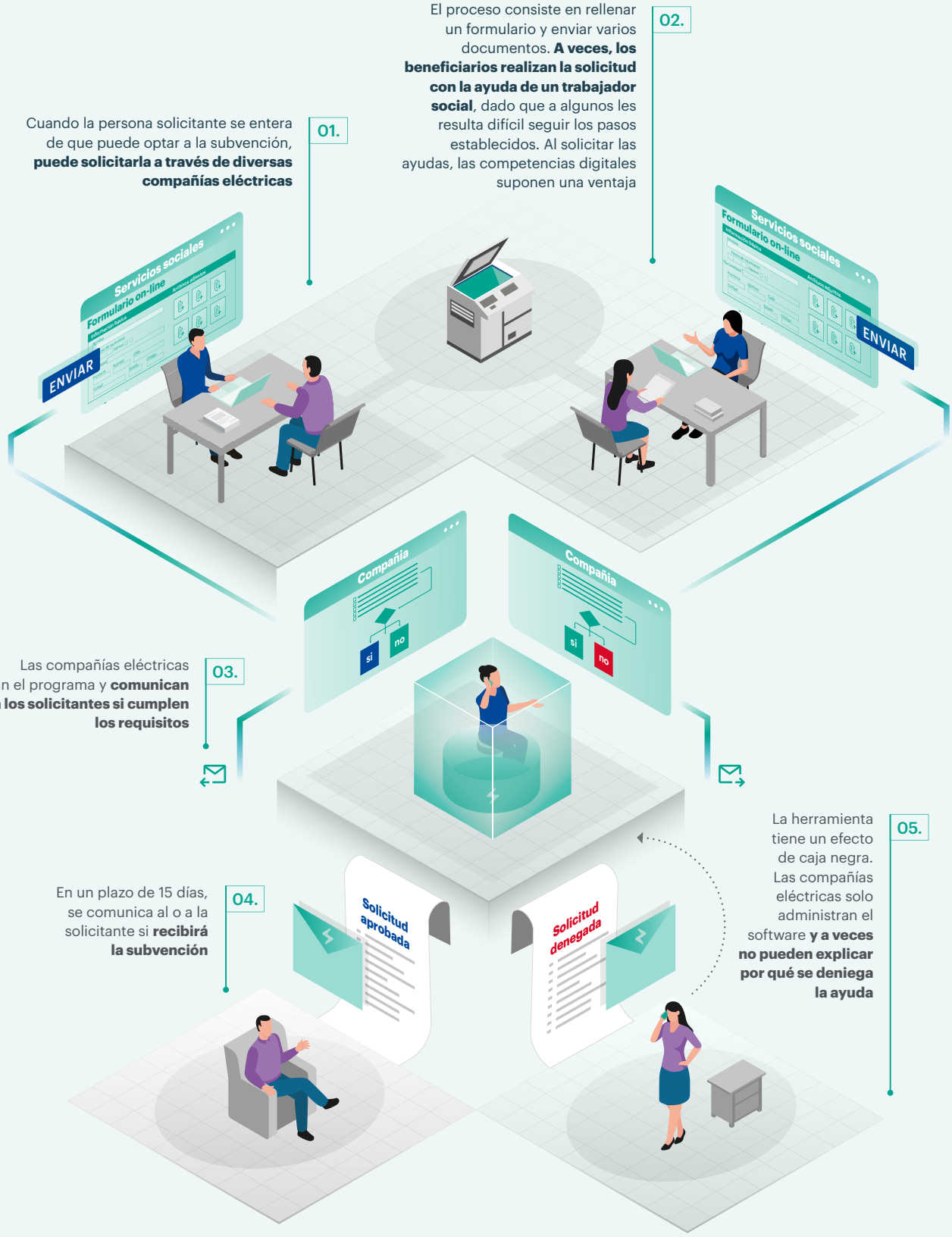


Figura 1. BOSCO.
Fuente de la imagen: Digital Future Society.

Discusión

El estado de bienestar digital y su impacto

Una de las mayores críticas a la reforma del 2017 (el cambio de los requisitos para recibir la ayuda y la implementación del programa BOSCO) es el elevado número de personas a las que se les deniega la ayuda. Según las estadísticas oficiales, recibieron la subvención 1,1 millones de consumidores, es decir, un tercio de los beneficiarios que había en el 2010 (3 millones), antes de que se promulgara la reforma (Comisión Nacional de los Mercados y la Competencia 2014, 2022).

Como ya se ha mencionado, esta diferencia se debe en parte al cambio en los criterios de acceso aplicado por el Gobierno. Pero Civio afirma que hay muchas otras razones por las que se está rechazando a personas con bajos ingresos. Por un lado, una parte considerable de la población que tiene derecho al bono social lo desconoce, o bien el sistema de solicitud en línea les resulta demasiado difícil o incomprensible (como en el caso de Mercedes). Por otro, el software rechaza solicitudes erróneamente.

Casos como el de BOSCO y su impacto en las poblaciones vulnerables forman parte de una tendencia más amplia de digitalización del estado de bienestar. Uno de los más críticos con esta digitalización es Philip Alston, ex relator especial de la ONU sobre la extrema pobreza y los derechos humanos. En su informe del 2019, expone la iniciativa de digitalizar los sistemas de bienestar, aparentemente benigna, como una oportunidad para reducir el presupuesto dedicado al bienestar, lo que implica "un estrechamiento del conjunto de beneficiarios, la eliminación de algunos servicios... y **una involución absoluta de la idea tradicional de que el Estado debe rendir cuentas al individuo**" (Alston 2019).

Una caja negra dentro de otra

Tal y como se ha mencionado, las compañías eléctricas (entidades privadas) administran el programa y comunican a los solicitantes si se les concede la ayuda. Esto añade una capa de complejidad para los usuarios finales —los beneficiarios—, que no están en contacto directo con la Administración pública. Una responsabilidad tan difusa supone un obstáculo más a la hora de impugnar la decisión automatizada. Las compañías eléctricas no tienen ningún interés en asegurarse de que la subvención se atribuya correctamente: se limitan a seguir órdenes y a aplicar las normas de selección establecidas por el programa informático. Si un usuario final cumple los requisitos para acceder a la ayuda y quiere saber por qué se le ha excluido, no hay a quien preguntar ni ningún modo de averiguar el motivo.

Este efecto genera una caja negra dentro de otra. Las compañías de electricidad no tienen constancia del proceso lógico que lleva a rechazar una subvención: pueden saber si falta documentación, pero no pueden ir más allá para entender si la ayuda se ha denegado por un error administrativo o porque el solicitante no cumple las condiciones exigidas. Y, dado que no son propietarias del software, sino que solo lo administran, no tienen ningún modo de entender a qué se deben las denegaciones de la ayuda. Tampoco los ciudadanos tienen acceso directo a la Administración para impugnarlo.

La lucha por la transparencia

Aunque el software BOSCO fue desarrollado internamente por la Administración española, los algoritmos que hay detrás de este tipo de sistemas suelen comprarse a empresas del sector privado y, por lo tanto, son privados (AI Now Institute 2018). La mayoría de las veces, el código fuente no se divulga, al estar protegido por "el secreto comercial". En este caso, la reticencia del Gobierno a publicar el código fuente amparándose en los derechos de propiedad intelectual resulta aún más sorprendente, dado que es la Administración pública quien desarrolló la herramienta.

Tanto si estos sistemas son privados como si no, esa renuencia es, por desgracia, un obstáculo habitual para los agentes externos como Civio. También es una traba importante para poder evaluar el sistema y detectar errores en el diseño o la implementación del software.

Sin la plataforma Civio, las numerosas familias a las que se les denegó la ayuda no habrían sabido que la denegación podía deberse a un error en el programa informático. La iniciativa de Civio se basa en la Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno, al argumentar que los beneficiarios tienen derecho a entender cómo se utilizan las decisiones automatizadas para determinar el acceso a los servicios. Algunos incluso sostienen que el Reglamento General de Protección de Datos (RGPD) exige que la información se presente de forma que los destinatarios puedan entenderla (Selbst y Powles 2017).

Cuando un algoritmo del sector público llega a los tribunales (casos de otros países)

La tendencia general de digitalizar los sistemas de asistencia social puede adoptar formas más sofisticadas, como en las herramientas para detectar fraudes relacionados con las prestaciones. Una de esas herramientas es el Servicio de señalización de fraudes (FSV) de los Países Bajos, un sistema de elaboración de perfiles de riesgo tristemente célebre, que a principios del 2021 provocó la dimisión del Gobierno de Rutte por infringir el RGPD.

El Gobierno holandés llevaba años aplicando un algoritmo diseñado para crear perfiles de riesgo de los residentes con mayor probabilidad de cometer fraude al recibir las prestaciones por hijos a cargo. El escándalo comenzó en el año 2012, y se supo que nada menos que 26.000 progenitores inocentes habían sido identificados como defraudadores y que a muchas familias se les había exigido devolver grandes sumas de dinero, que ascendían a decenas de miles de euros. Para agravar el escándalo, 11.000 familias de doble nacionalidad fueron objeto de un escrutinio especial, ya que el algoritmo utilizó ilegalmente datos como la nacionalidad de los solicitantes (Holligan 2021).

En los Países Bajos, la automatización ha cambiado drásticamente la forma de trabajar de los funcionarios. Marlies van Eck, profesora asistente de la Universidad Radboud de Nimega, asevera que, antes de que se automatizaran los pagos de prestaciones, estos se sometían a un exhaustivo proceso de revisión. El creciente uso de ADMS, en cierto modo, ha sustituido al criterio del personal de asistencia (Geiger 2021; NL Times 2021).

Pero la falta de transparencia persiste a nivel institucional. A principios del 2021, el organismo tributario holandés envió una carta a 60.000 personas comunicándoles el cierre del sistema FSV por haber incumplido la legislación en materia de protección de datos. Algunos de los que recibieron la carta no sabían que estaban en la lista ni por qué.

Instrumentos de evaluación de riesgos: nota explicativa

Antes de profundizar en los estudios de caso segundo y tercero, en esta sección se explica brevemente qué son los instrumentos de evaluación de riesgos, para qué se usan dentro de la justicia penal y cómo se utilizan. Este contexto aporta la información necesaria para entender las cuestiones abordadas en ambos casos.

¿Qué son los instrumentos de evaluación de riesgos?

Un instrumento de evaluación de riesgos (RAI, de *risk assessment instrument*) es un tipo de herramienta que se usa para predecir riesgos futuros. En el ámbito de la justicia penal, los RAI se usan para predecir el riesgo de que un acusado cometa en el futuro una conducta indebida, y se utilizan habitualmente para tomar decisiones judiciales previas a los juicios. Este uso de los RAI se extendió en los años setenta en países como Estados Unidos y el Canadá, donde las evaluaciones clínicas basadas en el criterio de los profesionales se consideraban muy subjetivas.

Dichas evaluaciones, fundamentadas en la experiencia clínica y que formalmente se denominan *evaluaciones clínicas no estructuradas*, recaen en expertos, como psiquiatras o psicólogos, y ayudan a los jueces a dictar sentencias y tomar otras decisiones antes de un juicio. Los RAI, a diferencia de las evaluaciones clínicas, proporcionan una predicción estructurada y basada en pruebas, y se incorporaron para reducir el peso del criterio personal y aumentar la objetividad (Marco Francia 2020).

Entre los RAI más comunes se encuentran los instrumentos actuariales de evaluación de riesgos (ARAI). Estas herramientas se basan en la evaluación estadística de diversos factores de riesgo predefinidos, y los partidarios de emplearlas defienden su potencial de tomar decisiones más coherentes, correctas y transparentes. Pero la objetividad que prometen ofrecer estos mecanismos de evaluación sigue siendo tema de debate (Heilbrun et al. 1999).

Quienes cuestionan su uso muestran preocupación por la exactitud y corrección de sus predicciones, su falta de individualización (algo que sí proporcionan las evaluaciones clínicas), sus sesgos y su falta de transparencia (Silver y Miller 2002). Aparte de generar preocupación sobre si son tan precisas, certeras y justas como los operadores humanos, estas herramientas son controvertidas porque han presentado sesgos de raza y de género.

Ese fue el caso del sistema de evaluación de reclusos Offender Assesment System (OASys) del Reino Unido (Angwin et al. 2016). Se descubrió que OASys, una herramienta comparable a la que se usa en el sistema de justicia penal español, RisCanvi, generaba predicciones diferentes en función de la raza, el género y la edad (Big Brother Watch 2020). Asimismo, herramientas parecidas, como COMPAS, de la compañía Northpointe, han sido objeto de controversia en Estados Unidos, dado su carácter privado y su falta de transparencia, al ampararse en el secreto comercial.

Las herramientas actuariales, como las que acabamos de mencionar, contienen *factores estáticos*, que incluyen la edad, la nacionalidad y los antecedentes, y *factores dinámicos*, de naturaleza socioeconómica y personal. Los indicadores de los factores estáticos no se pueden modificar, mientras que los factores dinámicos (como el abuso de sustancias) son potencialmente modificables.

La combinación de ambos métodos (las evaluaciones clínicas no estructuradas y las evaluaciones actuariales) se denomina *juicio profesional estructurado*. En él, los expertos clínicos utilizan como guía los indicadores estadísticos de riesgo y los emplean para tomar una decisión durante su análisis.

¿Cómo se evalúan las herramientas actuariales de evaluación de riesgos?

Las herramientas de evaluación de riesgos se evalúan en función de su precisión, calibración y discriminación. Si una herramienta es precisa, significa que identifica correctamente tanto los verdaderos positivos (presenta lo que se denomina **sensibilidad**) como los verdaderos negativos (presenta **especificidad**). Cuanto más preciso sea un sistema, menor será su porcentaje de errores; en el ámbito de la justicia penal una herramienta precisa predeciría correctamente si los reclusos reincidirán o no.

La calibración mide cuánto capta el riesgo absoluto una herramienta actuarial. Si está bien calibrada, predecirá la probabilidad de un resultado con la probabilidad observada. Si, por ejemplo, basándose en los datos históricos, se ha constatado que el 9,5 % de la población reclusa de Cataluña reincidirá, una herramienta bien calibrada evaluaría a la población actual en consonancia con la probabilidad observada.

La discriminación identifica el riesgo relativo. Es decir, mide lo bien que separa una herramienta a las personas de alto riesgo de las de bajo riesgo. Una forma habitual de medir la discriminación es utilizar el método del área bajo la curva (AUC). Si una herramienta tiene una puntuación de AUC aceptable (entre 0,7 y 0,8), significa que identifica a los reclusos de alto riesgo mejor que si se lanzara una moneda al aire. Si la puntuación de AUC es de 0,5, quiere decir que la herramienta no es capaz de detectar a los reclusos de alto riesgo mejor que si se lanzara una moneda al aire.

Según el *think tank* Urban Institute, la puntuación de AUC es una buena forma de determinar si una herramienta distingue eficazmente entre las personas de alto riesgo y las de bajo riesgo, pero no mide la probabilidad de reincidencia. Por ello "el método del AUC puede ser útil si se quiere usar la herramienta para determinar cómo dividir y asignar una cantidad fija de recursos en función de las prioridades" (Tiry y Kim 2021).

Caso nº 2: **RisCanvi**

Predicción del riesgo de reincidencia delictiva de los reclusos en Cataluña

El contexto

El verano del 2007, Francis Evrard, un hombre de 62 años encarcelado en Francia por violación en serie y pederastia, salió de prisión. Un mes después, reincidió al secuestrar y violar a un niño de 5 años (Savary 2009). En Cataluña, Alejandro Martínez Singul, delincuente sexual y violador en serie, salió de prisión poco después, causando gran indignación y preocupación en Barcelona, donde se temía que repitiera lo que había hecho Evrard unos meses antes.

Ese año, a raíz de la indignación reflejada en los medios de comunicación, se pidió a los responsables políticos un cambio en la legislación y se cuestionó la eficacia del sistema penitenciario. A fin de evitar casos como este, Montserrat Tura, la entonces consejera de Justicia de Cataluña, creó una comisión de investigación para proponer iniciativas que frenaran el ciclo de reincidencia.

Dicha comisión estaba integrada por diversos expertos, entre ellos, psicólogos, juristas, endocrinólogos, etc., y se centró especialmente en la violencia sexual y la reincidencia en delitos sexuales. De la comisión surgieron varias recomendaciones, con los siguientes resultados: se implantó la libertad vigilada y se encargó el diseño de un protocolo de evaluación de riesgos para valorar el riesgo de reincidencia, ahora denominado RisCanvi.

RisCanvi: los orígenes

Los instrumentos de evaluación de riesgos, como se ha mencionado, se utilizan cada vez más en todo el mundo para dictar sentencias, imponer penas de prisión y tomar decisiones sobre la libertad condicional. En el 2008, el Departamento de Justicia de Cataluña encargó al Grupo de Estudios Avanzados en Violencia, dirigido por Antonio Andrés Pueyo, catedrático de Psicología Criminal y de la Violencia, el diseño de una herramienta de evaluación actuarial para el contexto catalán.

Dicho Grupo tardó dos años en desarrollar un protocolo diseñado en estrecha colaboración con el Departamento de Justicia. Tuvieron acceso a cinco años (2003–2008) de datos internos sobre la población reclusa para comprender e identificar los factores de riesgo de los exreclusos que reincidieron posteriormente. Basándose en los datos de unos 600 presos, el grupo de investigación creó RisCanvi (cuyo nombre proviene de las palabras *riesgo* y *cambio* en catalán), que inicialmente se utilizó para predecir la probabilidad de que se dieran cuatro resultados. Desde su creación en el año 2010, la herramienta ha pasado por tres iteraciones y ahora calcula cinco resultados:

1. Violencia autodirigida: intentos de suicidio o lesiones autoinfligidas en el centro penitenciario
2. Violencia hacia otros reclusos o hacia el personal
3. Reincidencia violenta
4. Incumplimiento de la libertad condicional
5. Reincidencia general

Para predecir la probabilidad de cada uno, el protocolo emplea 43 factores de riesgo. En este tipo de herramienta, un factor de riesgo es una variable que ha demostrado estar muy correlacionada con la reincidencia. Para establecerlos, el grupo de investigación llevó a cabo una extensa revisión bibliográfica de otros métodos y protocolos, para comprender cómo se desarrollaron otras herramientas y qué factores de riesgo se tuvieron en cuenta.

¿Cómo se utiliza?

Cuando se registra a un recluso en el sistema penitenciario, primero pasa por una herramienta de cribado llamada RisCanvi-Screening (RisCanvi-S): con la que se realiza una evaluación a base de preguntas con dos opciones de respuesta (*sí* o *no*), para determinar el nivel de riesgo inicial de reincidencia. La herramienta de cribado consta de 10 factores, que se complementan con el informe y las entrevistas iniciales. Estos factores incluyen información como la edad a la que la persona participó por primera vez en un delito violento y si cuenta con apoyo familiar.[2] Esta primera herramienta se compone de factores estáticos que no pueden modificarse. La herramienta de cribado clasifica al recluso como de bajo o de alto riesgo. Si se determina que es de bajo riesgo, vuelve a someterse al cribado seis meses después.

Si el recluso se clasifica como de alto riesgo, se somete a RisCanvi-Complete (RisCanvi-C). Un equipo de profesionales multidisciplinar recoge datos sobre cada factor, junto con su historia clínica, observaciones y entrevistas. El equipo introduce esta información en la herramienta, que asigna al recluso una puntuación de riesgo. El resultado del algoritmo es solamente una clasificación con tres niveles, según la cual el riesgo puede ser bajo, medio o alto.

El equipo analiza la evaluación final para determinar el tipo de tratamiento que recibirá el interno. Se evalúa a los reclusos al menos cada seis meses, salvo en los casos de mala conducta en prisión, como la violencia contra otros presos, o en los de intento de suicidio. Si se produce un episodio de este tipo, se modifica la evaluación. Los niveles indicados por las predicciones también se incluyen en los informes que se envían a los fiscales y los jueces para que decidan o no aplicar medidas como la libertad condicional.

[2] Véase el anexo I para consultar todos los factores de riesgo de RisCanvi-S y RisCanvi-C.

RisCanvi

La herramienta se usa para valorar el riesgo potencial del recluso y la probabilidad de que se comporte de manera violenta dentro de la prisión o, en caso de salir de ella, el tipo de riesgo de reincidencia que presenta. Los resultados se consultan a lo largo de todo el "ciclo de vida" del recluso dentro del sistema penitenciario. Esto ayuda a la dirección a asignar recursos y a determinar el tratamiento posterior del interno.

RisCanvi-C emplea, como hemos dicho, 43 factores de riesgo, clasificados en cinco áreas principales. Estos son algunos de los factores que se usan en dicha herramienta de evaluación (Moreno Yuste 2015):

1. Antecedentes delictivos
    - Tipo de delito violento
    - Edad de comisión del primer delito
    - Comisión del delito bajo la influencia de drogas o alcohol
    - Delito con lesiones a la víctima
2. Antecedentes penitenciarios
    - Estancias anteriores en prisión
    - Conflictos en prisión
3. Antecedentes personales
    - Estancias de familiares en prisión
    - Nivel educativo
    - Situación laboral
4. Historia clínica
    - Consumo de drogas
    - Enfermedad mental grave
    - Comportamiento sexual promiscuo
5. Personalidad
    - Actitudes procriminales o valores antisociales
    - Impulsividad o inestabilidad emocional

La herramienta comprende estos 43 factores de riesgo, tanto estáticos como dinámicos, junto con otras cuatro variables que incluyen el sexo, la nacionalidad, la edad y el estado de la condena (si el recluso está a la espera de la sentencia definitiva o cumpliendo condena). Según Pueyo, estas variables tienen un peso significativo en el resultado final: por ejemplo, el hecho de que el recluso haya nacido en el mismo país tiene un peso diferente en el factor de riesgo de la salud mental porque, en general, los inmigrantes gozan de mejor salud mental en el sistema penitenciario catalán. Por otro lado, factores de riesgo como el de la integración social podrían tener más peso en el caso de los inmigrantes, dado que cuentan con menos recursos (Digital Future Society 2022a 00:09:43).

RisCanvi es una herramienta para evaluar el riesgo de reincidencia general y reincidencia violenta de los reclusos en Cataluña

Figura 2. RisCanvi.
Fuente de la imagen: Digital Future Society.

Discusión

Medir la eficacia de las herramientas de evaluación de riesgos

Según un estudio del 2015 realizado por el equipo de Manel Capdevila Capdevila, del Centro de Estudios Jurídicos y Formación Especializada (CEJFE), la precisión de este sistema al predecir la reincidencia de los reclusos era buena. De acuerdo con los estándares de este tipo de herramientas, el 77 % de los presos clasificados como de riesgo medio-alto reincidieron, y el 57 % de los no reincidentes fueron clasificados como de riesgo bajo por el algoritmo (Capdevila et al. 2015).

Sin embargo, un estudio posterior de la criminóloga Lucía Martínez Garay criticó la forma en que se interpretaron estos resultados, y afirmó que las conclusiones eran demasiado optimistas y confundían la sensibilidad de la herramienta con la precisión general de sus predicciones. Sostiene que los resultados de Capdevila, en realidad, indican que la herramienta tiene menos sensibilidad: solo en el 18 % de los casos predijo correctamente la reincidencia en los reclusos de riesgo medio y alto; en cambio, predijo mejor los perfiles de riesgo bajo, con un alto grado de especificidad (95 %) (Martínez Garay 2016).

En cuanto a la discriminación, calculada con la puntuación de AUC, Martínez Garay concluye que RisCanvi se sitúa dentro de los parámetros de eficacia de otras herramientas actuariales de evaluación de riesgos. Estas mediciones son esenciales para comprender la manera óptima de usar la herramienta. Como se ha mencionado, cada parámetro puede tener una finalidad en cada punto del proceso de toma de decisiones, dado que la herramienta se utiliza en diferentes momentos a lo largo del "ciclo de vida" de los reclusos y la usan diferentes profesionales.

Los expertos subrayan que es necesario comprender con claridad las limitaciones de estas herramientas de evaluación de riesgos para poder utilizarlas de manera eficaz. En general, se recomienda usar las evaluaciones actuariales de riesgos como herramientas de asesoramiento y, según algunos expertos, es posible que su eficacia predictiva haya llegado al límite. Martínez Garay advierte principalmente sobre el uso de RisCanvi para tomar decisiones sobre la libertad o los derechos fundamentales de los reclusos (Ibid.).

Replicación de sesgos

Las herramientas de evaluación de riesgos como RisCanvi se han promovido como una manera de reducir los sesgos y proporcionar un análisis objetivo de los riesgos. Los defensores de esta herramienta destacan las ventajas de aplicar un protocolo definido dentro del sistema penitenciario. Afirman que confiar únicamente en las evaluaciones clínicas da lugar a información sesgada e inconexa. Pero muchos autores han criticado el argumento de que las puntuaciones de riesgo son herramientas más objetivas y precisas. Dado que estas herramientas se basan en datos históricos sobre la población reclusa, muchos advierten de que en realidad están reproduciendo sesgos.

Bernard Harcourt, abogado y teórico crítico, sostiene que las herramientas de evaluación de riesgos, al basarse en el pasado de un recluso (más concretamente, en sus antecedentes delictivos), emiten predicciones sobre las decisiones policiales, es decir, sobre la probabilidad de que lo detengan, y no sobre esa persona y su peligrosidad (Cardoso 2020).

Protocolos, objetividad y sesgos

En un estudio reciente en el que se exploraba el uso profesional de los RAI, los autores observaron que los usuarios profesionales de los RAI eran conscientes de los aspectos sobre- e infrarrepresentados en los datos, y de los posibles resultados discriminatorios que pueden derivarse de su uso (Portela et al. 2022). Algunos expertos argumentan que es imposible eliminar los sesgos y subrayan que, al utilizar estas herramientas, los profesionales deben estar preparados adecuadamente para identificar y contrarrestar dichos sesgos.

Según el investigador Manuel Portela, los profesionales que saben usar RisCanvi son conscientes de las limitaciones de esta herramienta. Su estudio demuestra que los usuarios ven, eso sí, la herramienta como un recurso útil en su trabajo. Creen que proporciona una gran cantidad de datos, lo que da una perspectiva más "objetiva" que el criterio de los trabajadores sociales por sí solo.

Una vez más, el contexto es importante para comprender el riesgo potencial de amplificar los sesgos. Siempre hay un margen de error en el proceso de toma de decisiones, tanto en las personas como en los algoritmos. Por este motivo, RisCanvi se somete a la evaluación de un grupo de trabajadores sociales: la decisión no recae en una sola persona.

El uso de RisCanvi, el valor de la información que proporciona y la objetividad de sus usuarios son objeto de controversia. RisCanvi se diseñó para utilizarse en centros penitenciarios y de reinserción, pero también emplean esta herramienta los jueces, por ejemplo, para determinar si un recluso debe ser puesto en libertad condicional (Garrett y Monahan 2019).

Defender los derechos de los reclusos

Según la abogada Núria Monfort, experta en derecho penal y miembro de IACTA —una cooperativa sin ánimo de lucro que considera el derecho como una herramienta de transformación social—, para quienes ejercen la abogacía, RisCanvi es algo misterioso (una caja negra): apenas se entiende cómo funciona. Afirma que, aunque ahora sabemos cuáles son los 43 factores de riesgo, aún queda mucho por saber sobre cómo se calcula el riesgo y la ponderación que se asigna a cada variable (Digital Future Society 2022b 00:17:20).

Lo que sí saben es que se utiliza para predecir el riesgo futuro de reincidencia, lo que según Monfort supone un cambio radical en la práctica del derecho: antes de que se usara RisCanvi, había un diálogo, pero al incluirlo en el sistema, hay poco margen para el debate. El factor de riesgo se presenta como una verdad científica, difícil de rebatir.

Según explica, "el RisCanvi, o su necesidad, forma parte de todo un cambio de paradigma que empezó con pasar a sociedades que intentan evitar el riesgo... Es el paradigma de la seguridad frente al de la libertad". Señala que el riesgo o la peligrosidad no es algo nuevo; lo nuevo es que ahora hay un algoritmo para medir esa peligrosidad (Ibid. 00:22:00).

RisCanvi adquiere importancia cuando los reclusos pueden obtener permisos de salida o de libertad condicional. Esto sucede en España con sus tres grados de reclusión, dependiendo del tipo de delito cometido y el nivel de peligrosidad del recluso. A los de tercer grado, de bajo riesgo, se les permite disfrutar de un régimen de semilibertad, y la herramienta se usa para ayudar a calcular dicho riesgo.

Monfort apunta que las sociedades deben comprender la influencia política e ideológica de estas herramientas, que distan mucho de ser objetivas y científicas, y que priorizan la eficiencia sobre los derechos humanos de los presos. Añade: "Tenemos una población penitenciaria enorme... y hay un número limitado de profesionales y han de sacar un número de resoluciones. Pues vamos a tener una máquina. Esto no tiene nada que ver con una función social, con una responsabilidad colectiva, con una conducta individual del sujeto... Tiene que ver con cómo mover esta mole hacia algún lugar" (Ibid. 00:22:25).

Según Monfort, cada cambio legislativo en el código penal se ha producido tras un caso de violencia al que han prestado mucha atención los medios de comunicación, generalmente casos de violencia sexual. Estos cambios se llevaron a cabo partiendo de una determinada percepción del riesgo social, centrándose en un tipo de delito que no es de los más habituales. Pero RisCanvi se aplica a todo tipo de delitos, y sobre todo a un segmento de la población que está en riesgo de exclusión y que no tiene nada que ver con el tipo de infracción para el que se desarrolló esta herramienta. "RisCanvi surgió para la previsión de delitos violentos —señala Monfort—, y ahora resulta que su foco son los delitos no violentos... que son los que dan riesgos de reincidencia más altos" (Ibid. 00:19:45).

Casos similares en otros países

Las herramientas de predicción de riesgos se utilizan mucho en multitud de países, pero existen pocos estudios sobre su alcance (qué herramientas se usan y dónde) y sobre sus consecuencias para los presos de esos lugares. Dichas herramientas, en concreto las que se emplean para predecir la reincidencia delictiva, han estado en el punto de mira por las críticas recibidas en países como Estados Unidos y el Reino Unido. Entre los numerosos riesgos que plantean estos sistemas está la discriminación, ya que llevan arraigados los sesgos de la sociedad.

En el Reino Unido existe un sistema de evaluación de reclusos, Offender Assessment System (OASys), comparable a RisCanvi y que se utiliza en las audiencias previas a los juicios, las sentencias y las decisiones sobre la libertad condicional.

Un análisis del servicio nacional de administración de reclusos del Reino Unido (National Offender Management Service), realizado en el año 2014, descubrió que el predictor de reincidencia y el predictor de violencia del sistema generaban predicciones diferentes en función de la raza, el género y la edad: la validez "fue mayor en las mujeres que habían cometido delitos que en el caso de los hombres; en las personas blancas que en las de origen étnico asiático, negro y mixto, y en las de mayor edad en comparación con las más jóvenes".

En COMPAS (siglas en inglés de "Elaboración de perfiles de reclusos para la aplicación de sanciones sustitutivas en gestión penitenciaria"), una herramienta similar utilizada en EE. UU., también se ha observado que "según las predicciones del sistema, los acusados negros presentan un riesgo de reincidencia mayor que el índice real, y sucede lo contrario con los acusados blancos" (Big Brother Watch 2020).

Caso nº 3: **VioGén**

Evaluación del riesgo de volver a ejercer violencia de género

El contexto

En el año 2004, el Parlamento español aprobó por unanimidad una ley histórica contra la violencia de género, la primera de este tipo en Europa, para "prevenir, sancionar y erradicar la violencia de género y asistir a sus víctimas" (Ministerio de Justicia 2009). Esta ley, con un enfoque integral, estableció directrices para su aplicación a través de determinados protocolos. Como señala la ley, "tales protocolos impulsarán las actividades de prevención, detección precoz e intervención continuada con la mujer sometida a violencia de género o en riesgo de padecerla". Esta ley sentó las bases del protocolo de evaluación de riesgos denominado VioGén (acrónimo de *violencia de género*), que diseñó en el 2007 un equipo de expertos dirigido por el científico y catedrático Antonio Andrés Pueyo, quien un año después crearía RisCanvi.

Al igual que en el caso de RisCanvi, lo que promovió la legislación para crear VioGén fue la indignación ciudadana tras la muerte de Ana Orantes, una mujer asesinada por su exmarido en 1997 tras aparecer en televisión hablando de la violencia que había sufrido durante y después de su matrimonio. En directo, explicó al público que había acudido varias veces a la policía, pero que no había servido de nada. Su brutal asesinato y la indignación pública que provocó hicieron que, a partir de entonces, tanto los políticos como la sociedad en general se tomaran en serio la violencia de género. Hasta su muerte, en España la violencia de género se consideraba un asunto privado, que solo ocurría en ocasiones puntuales, y prácticamente no existían leyes que protegieran a las víctimas (Minder 2020).

¿Qué es VioGén?

VioGén[3] es un sistema para estimar el riesgo de reincidencia en violencia de género. Lo utilizan las fuerzas policiales de España, lo que incluye a la Policía Nacional, la Guardia Civil y los cuerpos de policía locales, excepto los de Cataluña y el País Vasco. Aunque el sistema recae bajo la jurisdicción del Ministerio del Interior, lo usan también otros ministerios, como los de Igualdad, Servicios Sociales y Justicia.

¿Cómo funciona?

Cuando se denuncia a la policía un episodio de violencia de género (puede hacerlo un familiar, un testigo o un agente del orden, no necesariamente la víctima), se inicia un proceso administrativo en el que los agentes abren una investigación y rellenan un formulario online junto con la víctima. Dicho formulario recibe el nombre de *valoración policial del riesgo* o VPR 5.0. Consta de 35 indicadores divididos en cinco categorías, que puede marcar la policía si se aplican al caso de la víctima.

[3] Sistema de Seguimiento Integral en los casos de Violencia de Género.

Incluyen los siguientes:[4]

1. Historia de violencia en la relación de pareja, entre otros:
   - Intensidad de las agresiones sexuales y físicas producidas con anterioridad
   - Empleo de armas
   - Amenazas de muerte
   - Indicios de celos exagerados
2. Características del agresor, entre otros:
   - Situación laboral
   - Adicciones o abuso de tóxicos
   - Salud mental
3. Información sobre la víctima, entre otros:
   - Si está embarazada
   - Si es dependiente económicamente
4. Información relacionada con los menores:
   - Si la víctima tiene a su cargo menores de edad
   - Si la víctima teme por la integridad física de los menores
5. Factores agravantes:
   - Si la víctima ha expresado su intención de romper la relación
   - Si la víctima piensa que el agresor es capaz de agredirla con mucha violencia o incluso matarla

A algunos parámetros se les puede asignar un nivel de gravedad. Por ejemplo, al registrar la violencia previa, se indica la intensidad de las agresiones físicas: si existe maltrato físico, los agentes pueden identificar el grado de intensidad, desde arañazos y moratones hasta apuñalamientos o intentos de asfixiar a la víctima. Con esta lista, pueden evaluar la gravedad de la situación.

Hay cinco posibles resultados: riesgo no apreciado, bajo, medio, alto y extremo. Los agentes pueden modificar la puntuación según el riesgo que consideren que presenta la situación. Dependiendo del nivel de riesgo del primer informe (si se determina que el riesgo es medio, alto o muy alto), se rellenan periódicamente unos formularios denominados *valoración policial de la evolución del riesgo* (VPER), a fin de valorar si la situación ha mejorado o empeorado.

Los momentos en que se deben rellenar los VPER dependen del nivel de riesgo identificado en la primera evaluación. Si no hay incidentes que requieran antes la atención de la policía (por ejemplo, si no hay más denuncias), en los casos de riesgo extremo el seguimiento se realiza a las 72 horas; si son de alto riesgo, al cabo de 7 días; en las situaciones de riesgo medio, a los 30 días, y en las de bajo riesgo, a los 60 días.

El seguimiento de los casos puede abarcar diferentes zonas geográficas. En función del riesgo asignado a cada caso, los agentes elaboran lo que se llama un *plan de seguridad personalizado*. Si el caso es de bajo riesgo, se vigila levemente al agresor, y si el riesgo es mayor, es posible que tanto el domicilio de la víctima como el del agresor se sometan a una estrecha vigilancia (Catanzaro 2020; González Álvarez et al. 2018).

[4] Véase el anexo II para consultar todos los indicadores de riesgo de VioGén.

¿Qué ocurre con la puntuación de riesgo una vez presentada la denuncia?
En el sistema VioGén, se crea un caso correspondiente a la víctima que denuncia a su agresor ante la policía. Cada víctima puede tener varios casos, y lo mismo sucede con el agresor. Es decir, si un agresor tiene varias víctimas, se crea un caso para cada una de ellas. Por ello, VioGén contiene más casos que personas registradas en el sistema.

El primer paso del proceso es la denuncia. Una vez que la víctima presenta dicha denuncia, se detiene al agresor y se celebra una vista judicial en la que el juzgado decide si el caso es de violencia de género y, si lo es, qué medidas deben tomarse. Se tienen en cuenta los resultados de la evaluación del riesgo junto con otras informaciones proporcionadas por la policía, los testigos, etc.

El caso permanece activo en el sistema VioGén si requiere la atención de la policía una vez que se realiza la evaluación periódica mencionada más arriba. En el proceso de valoración del riesgo, si este disminuye hasta el punto de que los agentes consideran que ya no hay riesgo de que se vuelva a ejercer violencia, el caso pasa a estar inactivo. En cambio, si se produce otro incidente violento o se presenta otra denuncia, el caso se reactiva. Solo se pueden eliminar casos registrados en el sistema si hay una justificación legal para hacerlo.

¿Cómo se ha percibido la herramienta?
En el 2022 se cumplen 15 años desde la creación de esta herramienta. De acuerdo con el Ministerio del Interior, ha permitido registrar más de 700.000 casos (Ministerio del Interior 2022). Sin embargo, desde su implementación, el nivel de riesgo asignado a cada caso ha sido objeto de controversia. En el 2014, el diario *El Mundo* publicó un documento filtrado que revelaba que 14 de las 15 mujeres asesinadas ese año se habían clasificado como casos de bajo riesgo (Catanzaro 2020).

En la misma línea, en el 2018 se determinó que la situación de Itziar Prats era de riesgo bajo, después de que denunciara ante los agentes la violencia ejercida por su expareja. Denunció ante la policía y ante la juez que el agresor había amenazado con matar a las dos hijas de ambos, pero los agentes no consideraron que el riesgo fuera alto, dado que la violencia hacia los menores no era un indicador de la lista en aquel momento. En septiembre de ese mismo año, su exmarido asesinó a sus dos hijas (Álvarez 2019).

En marzo del 2019, VioGén se modificó para incluir indicadores como el riesgo de asesinato de las mujeres y los menores, indicadores que no se habían tenido en cuenta en el proceso desde su implementación en el 2007.

Desde que se empezó a usar esta herramienta, también se han realizado otras mejoras. Por ejemplo, las víctimas de violencia de género no suelen evaluar bien su propio riesgo de sufrir violencia en el futuro, dado que subestiman el maltrato que sufren. Por lo general, no se reconocen como víctimas. Sin embargo, si se les pregunta por sus hijos, sí admiten el peligro que afrontan. Este tipo de preguntas se han reformulado considerando la psicología de las víctimas y su percepción de la violencia (Fundación Eticas 2022).

¿Qué tal funciona la herramienta?

En un estudio publicado recientemente, en el 2020, los investigadores observaron que el rendimiento de VioGén es similar al de otras herramientas actuariales, con una sensibilidad (verdaderos positivos) del 84 %, una especificidad (verdaderos negativos) del 60 % y un valor de AUC de 0,80 (López-Ossorio et al. 2020).[5]

Considerando estos resultados, el equipo de investigadores afirma que la herramienta puede identificar los casos de riesgo extremo que acaban en homicidio. Sin embargo, Eticas pone de relieve que, en España, solo uno de cada cuatro casos de homicidio se produce después de que la víctima denuncie. De modo que, por desgracia, muchas de las víctimas de homicidio no entran en el proceso de valoración policial del riesgo y, por lo tanto, sus casos no se identifican como de alto riesgo (Fundación Eticas 2022).

La perspectiva de una jurista

Sonia Márquez trabaja como abogada para la Fundación Ana Bella, una organización de la sociedad civil de Sevilla que se centra en ayudar a supervivientes a luchar contra la violencia de género. En su experiencia, ha observado que tanto la herramienta como el resto del sistema podrían hacer mucho más por que participaran sus usuarias finales (las víctimas), aunque el propio protocolo otorgue determinados derechos a las víctimas que presentan denuncia, como el de que las acompañe un profesional del derecho (Digital Future Society 2022d 00:29:11).

En primer lugar, hay un gran desconocimiento sobre el sistema en sí: las víctimas no saben que se les asigna una puntuación de riesgo ni lo que eso implica. Esto revela un problema, ya que ni en el diseño ni en la evaluación de la herramienta se tiene en cuenta su impacto social en la población afectada.

Según una auditoría externa realizada por Eticas, el 48 % de las 31 mujeres entrevistadas valoraron negativamente su experiencia (Fundación Eticas 2022). Márquez da fe de que el proceso de las entrevistas puede resultar confuso por diversos motivos, como la forma de plantear las preguntas. Algunas tratan sobre la violencia en sí y otras sirven para determinar qué tipo de medida de protección se debe aplicar. Pero apenas se les explica a las víctimas por qué les hacen esas preguntas.

[5] Estos datos corresponden a la nueva escala de homicidios integrada en la herramienta, la escala H, diseñada para mejorar la predicción de los homicidios perpetrados por la pareja, junto con el resto de las funciones de VPR del sistema.

VioGén

En segundo lugar, las víctimas se encuentran en una posición de clara desventaja si no entienden los aspectos administrativos o burocráticos del proceso de denuncia. Como abogada, Márquez ofrece a sus clientas información útil para interactuar con el sistema. Es importante hacerlo, puesto que la puntuación de riesgo puede variar mucho en función de:

1. Quién toma nota de la denuncia: los agentes de policía tienen sus propios sesgos sobre la violencia de género.
2. La hora y el día en que se presenta la denuncia: algunas comisarías cuentan con personal especializado para recibir a las víctimas de violencia de género, pero esos agentes tienen un horario específico dentro de cada departamento.
3. Dónde se presenta la denuncia: los casos pueden producirse en un pueblo pequeño o en una capital, y eso influye en la manera de gestionarlos, ya que no todos los juzgados disponen de un equipo especializado en tratar con víctimas de violencia de género.

VioGén es un sistema que utilizan las fuerzas policiales en España para estimar el riesgo de reincidencia en violencia de género

Cuando se denuncia a la policía un episodio de violencia de género, se inicia un proceso administrativo en el que los agentes abren una investigación y rellenan un formulario online junto con la víctima. El informe resultante conlleva una clasificación del nivel de riesgo de sufrir violencia en el futuro: no apreciado, bajo, medio, alto o extremo

01.

El primer paso del proceso es la denuncia. A continuación, **se detiene al agresor y un juzgado decide si el caso es de violencia de género** y, si lo es, qué medidas deben tomarse. Se tienen en cuenta los resultados de la evaluación del riesgo, junto con otros datos proporcionados por la policía y los testigos

02.

En función del riesgo asignado a cada caso, **se elabora un plan de seguridad personalizado**

04.

Dependiendo del nivel de riesgo consignado en el primer informe, se rellenan con una u otra periodicidad unos formularios denominados VPER, **a fin de valorar si la situación ha mejorado o empeorado**

03.

Figura 3. VioGén.
Fuente de la imagen: Digital Future Society.

Discusión

Un paso en la dirección correcta

Gemma Galdón, fundadora de la Fundación Eticas, afirma que el cuestionario aplica correctamente los hallazgos del ámbito académico, dado que el diseño de las preguntas refleja la complejidad de las víctimas de violencia de género. VioGén también ha recibido buenas críticas por parte de diversas organizaciones de derechos de las víctimas, gracias a sus esfuerzos por recoger datos que permitan comprender mejor la realidad de las víctimas de violencia de género.

Galdón cree que ofrece muchas oportunidades porque **la automatización ha permitido recopilar y analizar información de un modo que no sería posible sin la tecnología**. Según Galdón, el algoritmo de VioGén, al ser estático, no es capaz de aprender de los resultados previos, como los casos identificados erróneamente como de bajo riesgo. Por ello, **el cuestionario no se puede ajustar en función de si el algoritmo logra o no determinar correctamente el riesgo anterior de las víctimas**.

Galdón ve un gran potencial en el análisis de los datos recogidos por la herramienta, que podría permitir a las partes implicadas entender mejor el problema. Por ejemplo, hay ciertas correlaciones que los datos han ayudado a comprender y que no se habían abordado en la literatura hasta ese momento, como el hecho de que el lugar donde trabaja la víctima condiciona en gran medida la probabilidad de que sufra malos tratos en el futuro.

Supervisión humana

Como se ha mencionado, las herramientas actuariales proporcionan una cierta estructura como herramientas de análisis estadístico para la toma de decisiones. En cuanto a VioGén, el sistema se ha diseñado para mejorar el trabajo diario de los agentes de policía y mitigar los posibles sesgos y variaciones individuales. La evaluación del riesgo se ha diseñado de forma que se pueda modificar, y la idea es que sirva de complemento a la experiencia profesional del policía.

Sin embargo, un estudio del 2014 reveló que la mayoría de los agentes (el 95 %) confiaban en el resultado automatizado (Catanzaro 2020). Dicho estudio cuestiona si los agentes de policía, como primera línea de contacto, son los expertos adecuados para valorar el riesgo de que se vuelva a ejercer violencia de género. Los especialistas **—psicólogos y médicos forenses— subrayan la limitación que supone que sean los policías quienes evalúen el riesgo**, puesto que no tienen los conocimientos profesionales necesarios para valorar los indicadores. Lo ideal sería que desempeñaran este rol profesionales que hubieran recibido formación específica.

Aparte de requerir conocimientos especializados que escapan a los que poseen los agentes de policía, como señala Sonia Márquez, hay una serie de factores culturales, de las instituciones y de las organizaciones que influyen en la manera en que se realizan estas evaluaciones. Según una auditoría externa de la Fundación Eticas, **de los 27.000 agentes que participaron en el proceso de vigilancia, solo 2.000 están especializados en violencia de género.**

Entre el personal que trata directamente con las víctimas, faltan profesionales especializados que entiendan sus necesidades, lo cual se suma a la desconfianza en el sistema inherente en las víctimas. El problema de la violencia de género va mucho más allá del alcance de este sistema. En el año 2020, durante la pandemia, hubo un descenso significativo de las denuncias por violencia de género, lo que plantea la cuestión de si las denuncias ante la policía son una medida de prevención adecuada (Álvarez 2021).

En los ADMS, es habitual que la supervisión humana se implemente *a posteriori*, para cumplir la normativa. La auditoría externa del sistema llevada a cabo por la Fundación Eticas plantea dudas sobre cómo se ha integrado en el sistema la supervisión humana. La falta de claridad acerca de cómo deberían usar la herramienta los agentes pone de manifiesto esta cuestión. Galdón explica que hay varias posibles interpretaciones de cómo pueden interactuar los funcionarios con las herramientas automatizadas: pueden entender que su trabajo consiste simplemente en manejar la herramienta, o bien usarla como parte de un proceso de toma de decisiones en el que se sienten responsables de la evaluación.

La perspectiva de las víctimas

Otra crítica a VioGén es que el diseño del sistema no tiene en cuenta la experiencia de las víctimas. Como se ha mencionado, muchas de ellas no saben que las está evaluando un algoritmo ni qué derechos tienen quienes sufren violencia. El cuestionario no se ha puesto a prueba con víctimas de violencia, ni se ha sometido el sistema a una evaluación por parte de grupos de trabajo para, por ejemplo, determinar si deberían añadirse otras preguntas.

De acuerdo con Galdón, "en algoritmos de impacto social, el trabajo con usuarios finales es imprescindible para asegurar la calidad de esos sistemas, y también la transparencia de esos sistemas. Las mujeres [registradas en VioGén] tienen derecho a saber cómo se está determinando qué medidas de protección policial y legal se les van a implementar" (Digital Future Society 2022b 00:23:32).

El caso de Itziar Prats pone de manifiesto diversos fallos en el diseño del sistema que podrían haberse evitado si se hubiera ajustado mejor con la perspectiva de las víctimas, aunque algunos problemas sí se han abordado, por ejemplo, incluyendo la variable de las amenazas de muerte. Los críticos de VioGén defienden que no deberían existir esos niveles de riesgo. Argumentan que no debería incluirse una categoría de "riesgo no apreciado" y que todos los casos deberían considerarse de riesgo, a fin de evitar falsos negativos. Hay casos en los que, como le ocurrió a Itziar Prats, las mujeres que piden ayuda sufren revictimización o ven cómo su caso se desestima porque la calidad de la información, o el testimonio de la víctima por sí solo, no se corresponde con las expectativas de los agentes o los jueces (Coronado 2022).

Otros modelos

Las evaluaciones del riesgo estandarizadas, que incluyen herramientas actuariales como VioGén o escalas de juicio profesional estructurado (JPE), llevan usándose desde la década de los noventa en Norteamérica. Algunos ejemplos son muy conocidos, como Danger Assessment, Spousal Assault Risk Assessment (SARA), Domestic Violence Screening Inventory y Ontario Domestic Assault Risk Assessment.[6] Estas herramientas se crearon en entornos clínicos y, más adelante, las adoptaron entidades de defensa de las víctimas. Ahora las utilizan los cuerpos de policía (Helmus y Bourgon 2011).

En Europa, se empezó a prestar atención a estos sistemas en la década de los 2000. La primera mención a ellos en el Parlamento Europeo tuvo lugar en el 2004, y dicha institución incentivó el desarrollo de herramientas de evaluación de riesgos como una medida adecuada para reducir la violencia de género. España y Suecia fueron pioneras en este aspecto, al integrar en sus marcos jurídicos la evaluación y la gestión de riesgos (European Institute for Gender Equality s. f.).

B-Safer, una adaptación de la herramienta estadounidense SARA, se implementó en Suecia en los años noventa. El Reino Unido utiliza DASH, una escala de JPE, desde el año 2009. Por su parte, Portugal adaptó el modelo español en el 2019 (La Vanguardia 2019). En el panorama internacional, hay un gran debate sobre si las evaluaciones del riesgo son una herramienta apropiada para predecir la violencia de género.

Algunas herramientas prevén la violencia mejor que otras. Aún hay mucho que aprender sobre la violencia de género en términos globales —ya que, en todo el mundo, la mayoría de los casos no se denuncian— y sobre los patrones de maltrato, como aquellos que no dejan rastro visible en forma de lesiones (ONU-Mujeres 2022).

En la literatura, los investigadores del ámbito académico coinciden en que es necesario ser prudentes al implementar herramientas de evaluación de riesgos, y subrayan que los usuarios deben ser conscientes de sus puntos fuertes y sus limitaciones para aplicarlas con responsabilidad. Esto incluye la importancia de contar con evaluadores (cualificados) que entiendan cómo interpretar los resultados. La falta o la imprecisión de los datos también afecta a la precisión de las herramientas y, por ello, comprender cómo se administran el cuestionario, la situación, etc., es un factor importante que deben tener en cuenta los evaluadores.

Aún queda mucho por averiguar sobre el rendimiento que ofrecen estas herramientas en poblaciones diversas, algo que han señalado los expertos, especialmente en relación con herramientas como SARA y VioGén (Helmus y Bourgon 2011). Esta, en particular, es una de las principales preocupaciones de la herramienta de IA para predecir la violencia de género con la que está realizando pruebas piloto la Policía de Queensland, en Australia (Smee 2021). Aunque dicha institución es consciente de que este modelo podría afectar de manera injusta a la comunidad indígena y a otras minorías, y ha tomado medidas para mitigar los sesgos en el sistema, las organizaciones de la sociedad civil de Queensland muestran sus recelos ante el riesgo de que estos modelos refuercen los sesgos de los datos históricos.

[6] Para obtener más información sobre estas herramientas, véase https://www.rma.scot/wp-content/uploads/2019/09/RATED-Collated-Intimate-Partner-Violence-Awaiting-Validation.pdf

Caso nº 4: **SALER**

Predicción de casos potenciales de corrupción en Valencia

El contexto

A principios de los 2000, la Comunidad Valenciana —la cuarta más poblada del Estado español— tenía mala reputación a causa de sus numerosos escándalos de corrupción. En el 2015, un artículo del diario *El País* cifró en 150 el número total de políticos electos imputados, desde diputados regionales hasta jefes de gobierno (Barbería 2015). Faltaba poco para que el Partido Popular perdiera el apoyo que le habían mostrado los votantes a lo largo de veinte años. Ese año eligieron a dos partidos de izquierdas que, con el apoyo de un tercero, formaron un Gobierno de coalición a partir del llamado "Pacto del Botànic" en las Cortes Valencianas.

Uno de los principales objetivos de la nueva coalición era recuperar la confianza y dejar atrás la mala fama que había adquirido el Gobierno regional a lo largo de las últimas décadas. Se crearon varias entidades específicas para luchar contra la corrupción, entre ellas, la Agencia Antifraude y la Conselleria de Transparencia, Responsabilidad Social, Participación y Cooperación (Conselleria de Transparencia).

La narrativa institucional consistió en no centrarse en los casos de corrupción previos, sino empezar de cero y adoptar nuevas formas de trabajar para restablecer la confianza de la ciudadanía en la Administración de la comunidad. En este nuevo marco, la recién creada Conselleria de Transparencia quiso establecer un sistema de alertas para detectar los casos de mala conducta y prevenir una serie de situaciones que podrían dar lugar a fraudes, corrupción o mala praxis.

El principal propósito del sistema de alertas era impedir que los casos anteriores de corrupción contaminaran al Gobierno que acababa de formarse. Era muy importante que el sistema estuviera diseñado para identificar posibles malas prácticas en tiempo real y acompañar a los funcionarios en el proceso de identificación y resolución de esos casos.

¿Qué es SALER?

El Sistema de Alertas Rápidas (SALER) de la Generalitat Valenciana se creó oficialmente en el 2018, cuando las Cortes Valencianas aprobaron una ley para regular su uso.[7] Se rige por la Inspección General de Servicios, el órgano de control interno de la Administración regional.

SALER es un **sistema informático basado en el análisis de datos y diseñado para prever casos potenciales de corrupción en la Administración pública**. Esta herramienta es la primera de este tipo que se desarrolla en España; sin embargo, no se ha implementado por completo en la Administración pública.

[7] La ley se puede consultar aquí: https://dogv.gva.es/datos/2018/11/08/pdf/2018_10294.pdf

Alfons Puncel Chornet, exsubsecretario de la Conselleria de Transparencia en la primera legislatura resultante del del Pacto del Botànic[8] (del 2015 al 2019), es el padre de la herramienta. Así cuenta el momento en que tuvo la idea de crearla:

> Una interventora delegada me comenta que le llegan una serie de papeles para devolver una fianza de una obra. En el proceso de control, sobre papel, descubre que la obra no se había acabado. Eso me hace saltar la idea de que, si todo estuviera digitalizado, los papeles no habrían llegado, se habría cortado en el primer paso (Digital Future Society 2022e 00:05:45).

La conceptualización y el diseño del sistema, técnicamente, comenzaron en el 2016 con otro nombre —SATAN[9]—, con la ayuda del Departamento de Informática de la Universidad Politécnica de Valencia (Cid 2018). Por entonces, en la Inspección General de Servicios faltaba personal y no había suficientes recursos para desarrollar esa herramienta. De modo que, en paralelo, la Conselleria de Transparencia trabajó para sentar sus bases y que el sistema pudiera usarse internamente. Esto requirió colaborar con otros departamentos, establecer un marco jurídico adecuado y contratar más personal.

¿Cómo funcionará esta herramienta en la práctica?

El sistema SALER está diseñado para usar datos de diversas fuentes de la Administración pública, entre ellas:

- Registros de contratos (procesos de licitación y ejecución)
- Información sobre pagos directos (proveedores, servicios contratados)
- Datos sobre subvenciones (organismo que las otorga, beneficiarios, facturas, etc.)

Además, SALER recopila información de bases de datos notariales, el Registro de la Propiedad y el Registro Mercantil.
Cruza los datos de estas fuentes para identificar conflictos en curso o de forma preventiva: por ejemplo, durante una licitación, puede emitir una alerta de conflicto de intereses. Señala los siguientes tipos de fraudes: conflictos de intereses, duplicidades en la financiación, pactos contra terceros y elusión o mala gestión de los procedimientos de contratación pública establecidos.

[8] Pacto que dio lugar a la formación del Gobierno de coalición de tres partidos de izquierdas que marcó un cambio tras los veinte años de gobierno del PP en la Comunidad Valenciana.

[9] Sistema de Alertas Tempranas Anticorrupción.

SALER

Cuando el sistema identifica uno de estos casos, se avisa a los inspectores de la Intervención General de Servicios. Un inspector abre una investigación, que puede derivar en una inspección más detallada o en el cierre del caso. Al profundizar en la investigación, el inspector puede determinar si se trata de un error o una negligencia, malas prácticas o un posible caso de fraude. Este sistema de alertas tempranas está diseñado para acompañar al funcionario mientras lleva a cabo este procedimiento. Las alertas no son visibles públicamente.

¿Cómo se ha diseñado el sistema?

El sistema se ha entrenado con patrones de casos reales de malas prácticas y realiza búsquedas en un conjunto de bases de datos para identificar el riesgo. Un equipo formado por miembros de la Inspección General de Servicios definió los indicadores y las consultas para generar los algoritmos.

¿En qué fase se encuentra SALER actualmente?

SALER está en una etapa inicial, en la que se le están incorporando varias bases de datos. Uno de los principales obstáculos a los que se ha enfrentado es la lentitud de la transformación digital del sector público. Para que este sistema informático funcione, necesita acceder a un sistema consolidado en el que se unifiquen los conjuntos de datos y que tenga suficiente información. Eso es difícil de conseguir, ya que no todos los departamentos y organismos disponen de datos de buena calidad, precisos y accesibles. El sistema emplea bases de datos de diversas procedencias, de modo que los datos no son sencillos de analizar, porque no todas las bases de datos están programadas de la misma manera. Además, es necesario preparar los datos para que el sistema pueda cotejarlos, lo cual requiere tiempo y esfuerzo.

Pero las dificultades técnicas no son lo único que está frenando la implementación de SALER: también el hecho de que quienes defendieron originalmente este sistema, el secretario y el subsecretario de la Conselleria de Transparencia, han cambiado de rol tras la primera legislatura resultante del Pacto del Botànic.

SALER es una herramienta para prever posibles casos de corrupción en la Administración pública valenciana

Figura 4. SALER.
Fuente de la imagen: Digital Future Society.

Discusión

La gestión de la confianza en SALER

Una de las principales preocupaciones relacionadas con los sistemas de alertas es la sospecha de que estas herramientas se usan para vigilar y castigar a sus usuarios (en este caso, los funcionarios públicos). Conscientes de que este método de inspección puede generar desconfianza y hacer que los usuarios invaliden el sistema, los promotores de SALER pusieron mucho énfasis en explicar la diferencia entre SALER y otras herramientas que sí sirven para vigilar y castigar. Y, aun así, se enfrentaron a una gran resistencia entre el personal.

La transparencia es fundamental para lograr confianza. Si se ofrece más transparencia, el sistema tendrá más partidarios en las instituciones. Pero los sistemas como SALER, cuyo objetivo tiene que ver con el interés público, como la lucha contra el fraude y la corrupción, no se someten a los mismos estándares de transparencia que otros sistemas.

Se está debatiendo en qué medida debería ser transparente SALER, con el argumento de que ofrecer demasiada transparencia podría ser contraproducente: los usuarios malintencionados podrían usar esa información para encontrar formas de cometer actos corruptos, lo que socavaría el sistema.

Cumplimiento de la normativa de protección de datos

Otra gran dificultad que presenta el uso de grandes cantidades de datos, algo necesario para el sistema SALER, es el cumplimiento de la normativa de protección de datos. El sistema empezó con mal pie: en el 2017, cuando se estaba debatiendo la legislación correspondiente a esta herramienta, la Agencia Española de Protección de Datos publicó un informe muy crítico con el sistema de alertas, en el que se determinaba que no cumplía con las disposiciones del artículo 23 del RGPD.

Para cumplir la normativa de protección de datos, los sistemas como SALER deben definir con claridad qué datos van a recoger y analizar, con qué fin y durante cuánto tiempo. La normativa puede verse como un freno a la innovación. Una de las principales ventajas de usar estos sistemas es el potencial que ofrece la IA a la hora de cruzar datos de distintas bases de datos y determinar los indicadores de riesgo. La legislación de protección de datos limita el potencial de herramientas como SALER, que, con aprendizaje automático, podrían permitir al sistema detectar malas prácticas no identificadas hasta entonces por el personal.

Sin embargo, el sistema tendría que justificar qué datos cruza y especificar a qué bases de datos tiene acceso y de qué derechos disponen las personas con respecto a esos datos. Los datos deben ser proporcionados y deben utilizarse en la menor medida posible para lograr el objetivo del sistema. Antes de usar el sistema, los objetivos especificados tienen que definir qué datos se deben analizar para evaluar el riesgo de corrupción.

Un concepto sólido con consecuencias inciertas

Oscar Capdeferro, experto en derecho administrativo, alaba esta herramienta y la considera una medida eficaz contra la corrupción (Capdeferro 2018). Desde una perspectiva conceptual, SALER se ajusta a las recomendaciones respaldadas internacionalmente, sobre todo en cuanto a cómo el sistema de alertas se basa en análisis de riesgos anteriores. Otro factor de SALER que cabe mencionar es que se ha diseñado para coordinarse con otras medidas anticorrupción. La Inspección de Servicios continúa haciendo un seguimiento de las sospechas y, si se confirma un caso de corrupción, existe una base jurídica que permite a la Intervención General de Servicios imponer sanciones.

La ley delimita el uso de la herramienta y establece garantías y derechos para los usuarios y usuarios finales del sistema y el organismo responsable de la herramienta. También exige que se actualice periódicamente para velar por la eficacia del sistema de alertas.

No obstante, aún queda mucho por saber acerca del funcionamiento real de esta herramienta en la práctica. La forma de abordar los casos puede provocar una serie de resultados, algunos de ellos inesperados. El uso de la herramienta por parte de los inspectores y su impacto para los funcionarios tienen que ver tanto con el diseño de la herramienta como con factores relativos a las organizaciones e instituciones. Una posible aplicación de SALER que sugiere Capdeferro es ampliar el alcance de este sistema a otras Administraciones locales, puesto que la corrupción suele concentrarse más a nivel local.

Otras herramientas utilizadas

Cada vez más, los Gobiernos ven estas herramientas, diseñadas para detectar el fraude y la corrupción a partir de datos, como una solución eficaz y eficiente para controlar el uso de fondos públicos. Algunas de ellas han suscitado interés en todo el mundo, como System Risk Indication (SyRI), un sistema diseñado por el Gobierno de los Países Bajos para detectar el fraude entre los beneficiarios de prestaciones. Al implementar dicha herramienta surgieron tensiones y, en el 2018, varias organizaciones de la sociedad civil exigieron la suspensión de SyRi, alegando que el sistema infringía el derecho a la privacidad de las personas y no contaba con suficientes medidas de protección. Más tarde, en el 2020, el tribunal de primera instancia de La Haya decidió que el sistema violaba el derecho a la privacidad y que el uso de datos era desproporcionado con respecto a su finalidad de interés público: detectar el fraude en las prestaciones.[10]

A mayor escala, la Unión Europea también ha considerado necesario frenar la corrupción asociada a la financiación. De acuerdo con un informe elaborado en el 2019 por el Tribunal de Cuentas Europeo, se estima que cada año se sustraen 390 millones de euros de fondos estructurales, y esta estimación se basa únicamente en los casos detectados (OCDE 2019).

[10] Sentencia del tribunal de primera instancia de La Haya del 2020: https://uitspraken.rechtspraak.nl/inziendocument?id=ECLI:NL:RBDHA:2020:1878.

La UE emplea desde el año 2014 ARACHNE, una herramienta de valoración de riesgos para ayudar a las autoridades a detectar fraudes relacionados con el Fondo Social Europeo y el Fondo Europeo de Desarrollo Regional (Comisión Europea s. f.a). A partir de un conjunto de indicadores de riesgo, ARACHNE identifica el riesgo de fraude e irregularidades en las fases de aprobación e implementación de los proyectos.

De momento hay unos 20 Estados miembros que utilizan esta herramienta de forma voluntaria, aunque, a largo plazo, el plan es que sea de uso obligatorio. El Gobierno español, por ejemplo, se ha comprometido a usar ARACHNE en la gestión de los proyectos financiados por la UE (EuroEFE 2021).

De cara al futuro

En el actual panorama pospandémico, los Gobiernos de todo el mundo se están enfrentando a problemas cada vez más complejos, y las dificultades para abordarlos se irán incrementando si no se cuenta con la ayuda de las tecnologías digitales. Los sistemas de IA permiten a los funcionarios optimizar los servicios, tomar decisiones basadas en datos, prever riesgos y asignar recursos.

No obstante, el uso de la IA en el sector público ha empezado a atraer la atención de los medios de comunicación, las organizaciones de la sociedad civil y el público en general por su potencial de perjudicar a los ciudadanos. En los estudios de caso se observa que los daños que pueden causar dichos sistemas son palpables y que, si no se aborda esta cuestión, pueden agravar las desigualdades. A medida que las sociedades toman conciencia de los riesgos, es importante que comprendan la situación de forma global.

El objetivo de este informe es paliar la brecha de conocimientos existente entre el público en general, los expertos y los tecnólogos. También pretende aunar conocimientos entre las comunidades de habla hispana e inglesa. Aparte de mitigar la escasez de publicaciones que hablan de estos sistemas desde una perspectiva no técnica, en ambos idiomas, es necesario poner en común lo que ocurre en España, aprender de ello y contribuir a ampliar la información generada sobre este tema.

Deben considerarse cuatro cuestiones principales extraídas de los estudios de caso.

1. **Transparencia.** Sin transparencia no se puede realizar una evaluación adecuada del sistema, lo que conlleva que, si se producen daños algorítmicos, como una discriminación o un trato injusto por usar el algoritmo, no es posible solucionarlo.

   Al hablar de transparencia y sistemas algorítmicos, se debe diferenciar entre la falta de transparencia del propio sistema (que puede ser un obstáculo para que los funcionarios entiendan cómo ha tomado una determinada decisión el algoritmo) y la transparencia de los actores, que, por muchos motivos, como la legislación en materia de privacidad de datos, tal vez no quieran revelar la información. Por ejemplo, la transparencia que reclama Civio al Gobierno de España es necesaria para saber si el software cumple con la ley.

2. **Supervisión humana y responsabilidad.** A menudo, los sistemas de IA se implementan para ayudar a las personas a tomar decisiones, más que para sustituirlas. Sin embargo, en la práctica, la supervisión humana adopta muchas formas y puede ser heterogénea o casi simbólica. Tanto las herramientas de evaluación de riesgos como el sistema de alertas SALER se han diseñado para contextos concretos, en que los especialistas usan la información que les proporciona el algoritmo junto con otros factores, a la hora de tomar decisiones. En el caso de SALER, la alerta es pues solo el primer paso para valorar los posibles casos de corrupción. Sin embargo, el uso de la herramienta VioGén por parte de la policía pone en cuestión la eficacia de la supervisión humana, dado que

la mayoría de los agentes no modifican la evaluación del riesgo. Aunque los expertos recomiendan que las herramientas de evaluación de riesgos como VioGén estén en manos de profesionales especializados, también se puede argumentar que la finalidad de esas herramientas es proporcionar a la policía un protocolo estructurado.

Aun así, se están empleando sistemas que liberan expresamente a los funcionarios de una carga administrativa, como demuestra el caso de BOSCO. La implementación de este software supuso un cambio en un proceso interno: antes eran los funcionarios quienes tomaban la decisión sobre cada solicitud, pero ahora se ha transferido a las compañías eléctricas, que no pueden acceder a toda la información sobre los motivos de rechazo de las solicitudes. La complejidad de la supervisión humana de estos sistemas pone de relieve que es necesario entender mejor cómo funcionan en la práctica estas herramientas, y si ayudan a decidir o simplemente liberan del peso de las decisiones a las personas encargadas.

3. **Discriminación.** En tres de los cuatro casos analizados —BOSCO, VioGén y RisCanvi— se observa un **impacto social evidente en un segmento vulnerable de la población**. Como se ha comentado en el caso de BOSCO, hay poca información disponible sobre el software en sí, lo que impide averiguar si la causa de exclusión de los beneficiarios es el algoritmo o un error de software. En cualquier caso, se podrían aplicar varias mejoras significativas al proceso de solicitud, considerando que la carga administrativa no garantiza que esta ayuda gubernamental llegue a todos los que la necesitan. Las herramientas RisCanvi y VioGén parten de un historial de datos estadísticos. Por ejemplo, RisCanvi trata datos delicados que están ligados de manera inextricable a grupos tradicionalmente discriminados en el sistema jurídico. Otras herramientas empleadas en otros países también han mostrado riesgos de discriminación. Es necesario investigar más para entender qué sesgos pueden reforzar los algoritmos.

4. **Inclusividad.** Las organizaciones de la sociedad civil son fundamentales para concienciar sobre estas herramientas y ayudar a quienes se enfrentan a ellas a entender mejor cómo interactuar con estos sistemas. Su labor pone de manifiesto la necesidad de incluir a los usuarios finales en el diseño y la implementación de las herramientas, para que sean inclusivas y a fin de prevenir la discriminación, algo que, por desgracia, raramente se tiene en cuenta al desarrollarlas.

   Un ejemplo llamativo es el de VioGén, que en su diseño inicial excluía a los menores de la evaluación del riesgo. Esto podría haberse evitado si los creadores del sistema hubieran tenido en cuenta la perspectiva de las personas que se registran en el sistema o lo usan cada día.

   Una comunicación clara con todas las partes interesadas, así como el codiseño, son formas de generar confianza en estos sistemas y contribuir a que evolucionen hacia una mayor inclusividad. En el caso de SALER, la comunicación era clave desde el principio, para sentar sus bases y que los diversos departamentos de la Generalitat Valenciana comprendieran el propósito de esta herramienta. Si los funcionarios creen que SALER se usa con fines de vigilancia, eso puede hacer que se opongan al sistema y, con ello, anulen su utilidad.

Por último, es importante mencionar que tanto en España como en la Unión Europea se están tomando medidas para mitigar los daños potenciales que pueden presentar los sistemas algorítmicos, abordando los aspectos que se han mencionado más arriba: **transparencia, responsabilidad, supervisión humana** y **discriminación.**

Eso se refleja, sobre todo, en la propuesta de Ley de IA que está sometiendo a deliberación la Comisión Europea; es una iniciativa pionera en todo el mundo para establecer una regulación horizontal de los sistemas de IA. Anticipándose a dicha normativa, España está en proceso de crear la Agencia Española de Supervisión de la Inteligencia Artificial —la primera de su clase en la Unión Europea— para supervisar el uso de la IA. Otro ejemplo de este tipo de iniciativas es el Observatorio de Algoritmos con Impacto Social creado por la Fundación Eticas. Este proyecto consiste en un buscador de algoritmos con impacto social que ofrece información sobre ellos, como su uso, su ámbito de aplicación y sus objetivos.

Aunque estos pasos van en la dirección correcta, sigue habiendo muchas preguntas sin responder sobre si esas medidas son eficaces a la hora de mitigar los riesgos que representa la IA para los derechos fundamentales. La incertidumbre deja claro que es necesaria una coordinación entre los diferentes actores y partes interesadas, para asegurarse de que estos sistemas no se conviertan en herramientas que perjudiquen a los más vulnerables, perpetuando los sesgos y excluyéndolos de su diseño. Comprender el impacto social de estas herramientas es la única forma de que las partes implicadas entiendan cómo pueden estos sistemas proteger los derechos fundamentales, y se cercioren de que dichos mecanismos no sean un obstáculo para las mismas instituciones a las que deberían ayudar.

Anexos

Anexo I: Factores incluidos en RisCanvi-S y RisCanvi-C

RisCanvi-Screening:
RisCanvi incluye información básica sobre cada recluso:
su edad, género, estado civil, situación procesal-penitenciaria, régimen de vida penitenciaria, tipo de delito y relación con la víctima. Posteriormente aparecen 10 ítems: 1) edad del primer incidente violento o inicio de las conductas violentas; 2) violencia previa (al delito principal); 3) comportamiento penitenciario anterior (faltas graves o muy graves); 4) evasiones, fugas, quebrantamientos de condena; 5) problemas con el consumo de alcohol u otras drogas; 6) problemas de salud mental anterior (diagnósticos previos de trastornos, ira, inestabilidad emocional, impulsividad); 7) intentos o conductas de autolesión previos; 8) falta de soporte familiar y social, falta de una red relacional; 9) problemas de índole laboral/económica; 10) actitudes hostiles o valores antisociales.

RisCanvi Complete:
Factores criminales/penitenciarios: 1. Delito base violento; 2. Edad en el momento del delito base; 3. Intoxicación durante la realización del delito base; 4. Víctimas con lesiones; 5. Duración de la pena; 6. Tiempo ininterrumpido en prisión; 7. Historia de violencia; 8. Inicio de la actividad delictiva o violenta; 9. Incremento de frecuencia, gravedad y/o diversidad de los delitos; 10. Conflictos con otros internos; 11. Incumplimiento de medidas judiciales; 12. Expedientes disciplinarios; 13. Evasiones o fugas; 14. Regresión de grado; 15. Quebrantamiento de permisos.
Factores personales/sociofamiliares: 16. Desajuste infantil; 17. Distancia entre su residencia habitual y el centro; 18. Nivel educativo; 19. Problemas relacionados con la educación; 20. Falta de recursos económicos; 21. Ausencia de planes viables de futuro; 22. Antecedentes delictivos en la familia de origen; 23. Socialización problemática en familia de origen; 24. Falta de soporte familiar y social; 25. Amistades criminales/delincuentes; 26. Pertenencia a grupos sociales de riesgo; 27. Rol delictivo destacado; 28. Víctima de violencia de género (solo aplicable a mujeres); 29. Cargas familiares actuales.
Factores clínicos/de personalidad: 30. Abuso o dependencia de drogas; 31. Abuso o dependencia del alcohol; 32. Trastorno mental severo; 33. Comportamiento sexual promiscuo; 34. Respuesta limitada al tratamiento psicológico o psiquiátrico; 35. Trastorno de personalidad relacionado con la ira; 36. Pobre afrontamiento del estrés; 37. Intentos o conductas de autolesión; 38. Actitudes procriminales o valores antisociales; 39. Baja capacidad mental e inteligencia; 40. Temeridad; 41. Impulsividad, inestabilidad emocional; 42. Hostilidad; 43. Irresponsabilidad.

Anexo II

1. HISTORIA DE VIOLENCIA EN LA RELACIÓN DE PAREJA	RESPUESTAS		
Indicador 1: Violencia psicológica (vejaciones, insultos y humillaciones)	**Sí**	**No**	**N/S**
1.1 Intensidad de la violencia psicológica	Leve	Grave	Muy grave
Indicador 2: Violencia física	**Sí**	**No**	**N/S**
2.1 Intensidad de la violencia física	Leve	Grave	Muy grave
Indicador 3: Sexo forzado	**Sí**	**No**	**N/S**
3.1 Intensidad de la violencia sexual	Leve	Grave	Muy grave
Indicador 4: Empleo de armas u objetos contra la víctima	**Sí**	**No**	**N/S**
4.1 Arma blanca, 4.2 Arma de fuego, 4.3 Otros objetos	Leve	Grave	Muy grave
Indicador 5: Existencia de amenazas o planes dirigidos a causar daño a la víctima	**Sí**	**No**	**N/S**
5.1 Intensidad de las amenazas	Leve	Grave	Muy grave
5.2 Amenazas de suicidio del agresor	**Sí**	**No**	
5.3 Amenazas de muerte del agresor dirigidas a la víctima	**Sí**	**No**	
Indicador 6: En los últimos seis meses se registra un aumento de la escalada de agresiones o amenazas	**Sí**	**No**	**N/S**
2. CARACTERÍSTICAS DEL AGRESOR			
Indicador 7: En los últimos seis meses, el agresor muestra celos exagerados o sospechas de infidelidad	**Sí**	**No**	**N/S**
Indicador 8: En los últimos seis meses, el agresor muestra conductas de control	**Sí**	**No**	**N/S**
Indicador 9: En los últimos seis meses, el agresor muestra conductas de acoso	**Sí**	**No**	**N/S**
Indicador 10: Existencia de problemas en la vida del agresor en los últimos seis meses	**Sí**	**No**	
10.1 Problemas laborales o económicos	**Sí**	**No**	
10.2 Problemas con el sistema de justicia	**Sí**	**No**	**N/S**
Indicador 11: En el último año el agresor produce daños materiales	**Sí**	**No**	**N/S**
Indicador 12: En el último año se registran faltas de respeto a la autoridad o a sus agentes	**Sí**	**No**	**N/S**
Indicador 13: En el último año agrede físicamente a terceras personas y/o animales	**Sí**	**No**	**N/S**

Indicador 14: En el último año existen amenazas o desprecios a terceras personas	Sí	No	N/S
Indicador 15: Existen antecedentes penales y/o policiales del agresor	Sí	No	N/S
Indicador 16: Existen quebrantamientos previos o actuales (cautelares o penales)	Sí	No	N/S
Indicador 17: Existen antecedentes de agresiones físicas y/o sexuales	Sí	No	N/S
Indicador 18: Existen antecedentes de violencia de género sobre otra/s pareja/s	Sí	No	N/S
Indicador 19: Presenta problemas o un trastorno mental y/o psiquiátrico	Sí	No	N/S
Indicador 20: Presenta ideas o intentos de suicidio	Sí	No	N/S
Indicador 21: Presenta algún tipo de adicción o conductas de abuso de tóxicos (alcohol, drogas y fármacos)	Sí	No	N/S
Indicador 22: Presenta antecedentes familiares de violencia de género o doméstica	Sí	No	N/S
Indicador 23: El agresor tiene menos de 24 años	Sí	No	N/S
3. FACTORES DE RIESGO/VULNERABILIDAD DE LA VÍCTIMA			
Indicador 24: Existencia de algún tipo de discapacidad, enfermedad física o psíquica grave	Sí	No	N/S
Indicador 25: Víctima con ideas o intentos de suicidio	Sí	No	N/S
Indicador 26: Presenta algún tipo de adicción o conductas de abuso de tóxicos (alcohol, drogas y fármacos)	Sí	No	N/S
Indicador 27: Carece de apoyo familiar o social favorable	Sí	No	
Indicador 28: Víctima extranjera			
4. CIRCUNSTANCIAS RELACIONADAS CON LOS MENORES			
Indicador 29: La víctima tiene a su cargo menores de edad	Sí	No	N/S
Indicador 30: Existencia de amenazas a la integridad física de los menores	Sí	No	N/S
Indicador 31: La víctima teme por la integridad de los menores	Sí	No	N/S
5. CIRCUSTANCIAS AGRAVANTES			
Indicador 32: La víctima ha denunciado a otros agresores en el pasado	Sí	No	N/S
Indicador 33: Se han registrado episodios de violencia lateral recíproca	Sí	No	N/S
Indicador 34: La víctima ha expresado al agresor su intención de romper la relación hace menos de seis meses	Sí	No	N/S
Indicador 35: La víctima piensa que el agresor es capaz de agredirla con mucha violencia o incluso matarla	Sí	No	N/S

Referencias

AI Now Institute. (2018). Litigating algorithms: challenging government use of algorithmic decision systems. [PDF] Disponible en: https://ainowinstitute.org/litigatingalgorithms.pdf (Consultado: 31-1-2023)

Alston, P. (2019). Informe del relator especial sobre la extrema pobreza y los derechos humanos. [online] Disponible en: https://www.ohchr.org/en/press-releases/2019/10/world-stumbling-zombie-digital-welfare-dystopia-warns-un-human-rights-expert (Consultado: 31-1-2023)

Álvarez, P. (2019). "Denuncié, me dijeron que no pasaría nada y mis hijas ya no están". El País. [online] Disponible en: https://elpais.com/sociedad/2019/03/23/actualidad/1553369290_857804.html (Consultado: 31-1-2023)

Álvarez, P. (2021). Más asesinatos machistas en 30 días que en los cuatro primeros meses de 2021. El País. [online] Disponible en: https://elpais.com/sociedad/2021-06-17/mas-asesinatos-machistas-en-30-dias-que-en-los-cuatro-primeros-meses-de-2021.html (Consultado: 31-1-2023)

Angwin, J., Larson, J., Mattu, S. y Kirchner, L. (2016). Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks. ProPublica. [online] https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (Consultado: 31-1-2023)

Barbería, J. L. (2015). Why Valencia is paying the high price of rampant political corruption. El País. [online] Disponible en: https://english.elpais.com/elpais/2015/05/06/inenglish/1430932717_848440.htm (Consultado: 31-1-2023)

Big Brother Watch. (2020). Big Brother Watch briefing on Algorithmic Decision-Making in the Criminal Justice System. [PDF] Disponible en: https://bigbrotherwatch.org.uk/wp-content/uploads/2020/02/Big-Brother-Watch-Briefing-on-Algorithmic-Decision-Making-in-the-Criminal-Justice-System-February-2020.pdf (Consultado: 31-1-2023)

Capdeferro Villagrasa, O. (2018). El análisis de riesgos como mecanismo central de un sistema efectivo de prevención de la corrupción. En particular, el sistema de alertas para la prevención de la corrupción basado en inteligencia artificial. Revista Internacional Transparencia e Integridad. [PDF] Disponible en: http://www.encuentros-multidisciplinares.org/revista-65/oscar_capdeferro-1.pdf (Consultado: 31-1-2023)

Capdevila Capdevila, M., Blanch Serentill, M., Ferrer Puig, M., Andrés Pueyo, A., Framis Ferrer, B., Comas López, N., Garrigós Bou, A., Boldú Pedro, A., Batlle Manonelles, A. y Mora Encinas, J. (2015). Tasa de reincidencia penitenciaria 2014. Centro de Estudios Jurídicos y Formación Especializada, Generalitat de Catalunya. [PDF] Disponible en: https://cejfe.gencat.cat/web/.content/home/recerca/cataleg/crono/2015/taxa_reincidencia_2014/tasa_reincidencia_2014_cast.pdf (Consultado: 31-1-2023)

Cardoso, T. (2020). Bias behind bars: A Globe investigation finds a prison system stacked against Black and Indigenous inmates. [online] Disponible en: https://www.theglobeandmail.com/canada/article-investigation-racial-bias-in-canadian-prison-risk-assessments/ (Consultado: 31-1-2023)

Catanzaro, M. (2020). In Spain, the VioGén algorithm attempts to forecast gender violence. AlgorithmWatch. [online] Disponible en: https://algorithmwatch.org/en/viogen-algorithm-gender-violence/ (Consultado: 31-1-2023)

Cedefop. (2021). Digital, greener and more resilient. Insights from Cedefop's European skills forecast. [PDF] Disponible en: http://data.europa.eu/doi/10.2801/154094/ (Consultado: 31-1-2023)

Cid, G. (2018). Ingenieros valencianos crean 'Satan', un 'software' para cazar corruptos: así funciona. Elconfidencial.com. [online] Disponible en: https://www.elconfidencial.com/tecnologia/2018-10-22/algoritmo-anticorrupcion-valencia-satan_1632428/ (Consultado: 31-1-2023)

Civio. (2019). Que se nos regule mediante código fuente o algoritmos secretos es algo que jamás debe permitirse en un Estado social, democrático y de Derecho. [online] Disponible en: https://civio.es/novedades/2019/07/02/que-se-nos-regule-mediante-codigo-fuente-o-algoritmos-secretos-es-algo-que-jamas-debe-permitirse-en-un-estado-social-democratico-y-de-derecho/ (Consultado: 31-1-2023)

Comisión Europea. (s.f.a). ARACHNE risk scoring tool. [online] Disponible en: https://ec.europa.eu/social/main.jsp?catId=325&intPageId=3587&langId.it (Consultado: 31-1-2023)

Comisión Europea. (s.f.b). SALER Rapid Alert System. [online] Disponible en: https://antifraud-knowledge-centre.ec.europa.eu/library-good-practices-and-case-studies/good-practices/saler-rapid-alert-system_en (Consultado: 5-1-2023)

Comisión Nacional de los Mercados y la Competencia. (2014). ¿Cuántos usuarios se benefician del bono social de electricidad en España? [online] Disponible en: https://blog.cnmc.es/2014/10/10/cuantos-usuarios-se-benefician-del-bono-social-de-electricidad-en-espana/ (Consultado: 5-1-2023)

Comisión Nacional de los Mercados y la Competencia. (2022). Estadística bono social. [online] Disponible en: https://data.cnmc.es/energia/energia-electrica/bono-social/estadistica-bono-social (Consultado: 5-1-2023)

Coronado, N. (2022). Una Guardia Civil de Viogen revictimiza a una madre en su llamada de socorro al denunciar que el maltratador no le entrega a su pequeña. La Hora Digital. [online] Disponible en: https://www.lahoradigital.com/noticia/32054/igualdad/una-guardia-civil-de-viogen-revictimiza-a-una-madre-en-su-llamada-de-socorro-al-denunciar-que-el-maltratador-no-le-entrega-a-su-pequena.aspx/ (Consultado: 31-1-2023)

Digital Future Society. (2020). Hacia la igualdad de género en el estado de bienestar digital. [PDF] Disponible en: https://digitalfuturesociety.com/app/uploads/2020/09/Hacia_igualdad_género_en_estado_bienestar_digital-1.pdf (Consultado: 31-1-2023)

Digital Future Society. (2021). Gobernanza y algoritmos. [PDF] Disponible en: https://digitalfuturesociety.com/es/report/governing-algorithms/ (Consultado: 31-1-2023)

Digital Future Society. (2022a). Capítulo 1. RisCanvi (I): el algoritmo de la cárcel. [pódcast] Disponible en: https://digitalfuturesociety.com/es/podcasts/capitulo-1-riscanvi-i-el-algoritmo-de-la-carcel/ (Consultado: 31-1-2023)

Digital Future Society. (2022b). Capítulo 2. RisCanvi (II): ¿Se puede predecir el próximo delito? [pódcast] Disponible en: https://digitalfuturesociety.com/es/podcasts/capitulo-2-riscanvi-ii-se-puede-predecir-el-proximo-delito/ (Consultado: 31-1-2023)

Digital Future Society. (2022c). Capítulo 3: BOSCO y el bono para pagar la luz. [pódcast] Disponible en: https://digitalfuturesociety.com/es/podcasts/capitulo-3-bosco-y-el-bono-para-pagar-la-luz/ (Consultado: 31-1-2023)

Digital Future Society. (2022d). Capítulo 4. VioGén, el software contra la violencia machista. [pódcast] Disponible en: https://digitalfuturesociety.com/es/podcasts/capitulo-4-viogen-el-software-contra-la-violencia-machista/ (Consultado: 31-1-2023)

Digital Future Society. (2022e). Capítulo 5. SALER y los algoritmos contra la corrupción. [pódcast] Disponible en: https://digitalfuturesociety.com/es/podcasts/chapter-5-saler-and-algorithms-against-corruption/ (Consultado: 31-1-2023)

EuroEFE. (2021). Spain approves €70 billion recovery plan to help transform economy. [online] Disponible en: https://www.euractiv.com/section/economy-jobs/news/spain-approves-e70-billion-recovery-plan-to-help-transform-economy/ (Consultado: 31-1-2023)

European Digital Rights. (2021). Civil society calls on the EU to put fundamental rights first in the AI Act. [online] Disponible en: https://edri.org/our-work/civil-society-calls-on-the-eu-to-put-fundamental-rights-first-in-the-ai-act/ (Consultado: 31-1-2023)

European Institute for Gender Equality. (s.f.). Risk assessment and risk management by police. [online] Disponible en: https://eige.europa.eu/gender-based-violence/risk-assessment-risk-management/areas-improvement/ (Consultado: 31-1-2023)

Fundación Eticas. (2022). Auditoría Externa del Sistema VioGén. [PDF] Disponible en: https://eticasfoundation.org/wp-content/uploads/2022/04/ETICAS-_-Auditori%CC%81a-Externa-del-sistema-VioGe%CC%81n-_-20220308.docx.pdf (Consultado: 31-1-2023)

Garrett, B. y Monahan, J. (2019). Judging Risk. [PDF] Disponible en: https://judicature.duke.edu/wp-content/uploads/2022/06/GARRETT-Summer-2019.pdf (Consultado: 31-1-2023)

Geiger, G. (2021). How a Discriminatory Algorithm Wrongly Accused Thousands of Families of Fraud. Vice. [online] Disponible en: https://www.vice.com/en/article/jgq35d/how-a-discriminatory-algorithm-wrongly-accused-thousands-of-families-of-fraud/ (Consultado: 31-1-2023)

González Álvarez, J. L., López Ossorio, J. J., Urruela Cortés, C. y Rodríguez Díaz, M. (2018). Integral Monitoring System in Cases of Gender Violence. VioGén System. Behavior & Law Journal. [PDF] Disponible en: https://behaviorandlawjournal.com/BLJ/article/download/56/65/299/ (Consultado: 31-1-2023)

Heilbrun, K., Dvoskin, J., Hart, S. y McNiel, D. (1999). Violence risk communication: Implications for research, policy, and practice. [online] Disponible en: https://www.tandfonline.com/doi/abs/10.1080/13698579908407009 (Consultado: 31-1-2023)

Helmus, L. y Bourgon, G. (2011). Taking Stock of 15 Years of Research on the Spousal Assault Risk Assessment Guide (SARA): A Critical Review. [online] Disponible en: https://www.tandfonline.com/doi/abs/10.1080/14999013.2010.551709 (Consultado: 31-1-2023)

Holligan, A. (2021). Dutch Rutte government resigns over child welfare fraud scandal. BBC. [online] Disponible en: https://www.bbc.com/news/world-europe-55674146/ (Consultado: 31-1-2023)

Kayser-Bril, N. (2019). Spain: Legal fight over an algorithm's code. [online] Disponible en: https://algorithmwatch.org/en/spain-legal-fight-over-an-algorithms-code/ (Consultado: 31-1-2023)

La Vanguardia. (2019). Portugal reforzará la protección a las víctimas de violencia machista. [online] Disponible en: https://www.lavanguardia.com/politica/20190207/46286941071/portugal-reforzara-la-proteccion-a-las-victimas-de-violencia-machista.html (Consultado: 31-1-2023)

López-Ossorio, J. J., González-Álvarez, J. L., Loinaz, I. y Martínez Martínez, A. (2020). Intimate partner homicide risk assessment by police in Spain: The dual protocol VPR5.0-H. [online] Disponible en: https://journals.copmadrid.org/pi/art/pi2020a16 (Consultado: 31-1-2023)

Marco Francia, M. P. (2020). Evaluando la peligrosidad criminal. [PDF] Disponible en: https://www.researchgate.net/publication/346734133/ (Consultado: 31-1-2023)

Martínez Garay, L. (2016). Errores conceptuales en la estimación de riesgo de reincidencia. Revista Española de Investigación Criminológica, 14, 1-31. [PDF] Disponible en: https://reic.criminologia.net/index.php/journal/article/view/97/94/ (Consultado: 31-1-2023)

Mileusnic, M. (2022). Plan nacional de recuperación y resiliencia de España. Servicio de Estudios del Parlamento Europeo. [PDF] Disponible en: https://www.europarl.europa.eu/RegData/etudes/BRIE/2022/698878/EPRS_BRI(2022)698878_ES.pdf (Consultado: 31-1-2023)

Minder, R. (2020). Ana Orantes, la mujer cuyo asesinato atroz hizo que España cambiara sus leyes. The New York Times. [online] Disponible en: https://www.nytimes.com/es/2020/01/17/espanol/ana-orantes-times.html (Consultado: 31-1-2023)

Ministerio de Asuntos Económicos y Transición Digital. (s.f.). España Digital 2025. [PDF] Disponible en: https://portal.mineco.gob.es/RecursosArticulo/mineco/prensa/ficheros/noticias/2018/Agenda_Digital_2025.pdf (Consultado: 31-1-2023)

Ministerio de Justicia. (2009). Ley Orgánica de Medidas de Protección Integral contra la Violencia de Género. [PDF] Disponible en: https://violenciagenero.igualdad.gob.es/definicion/pdf/LEY_ORGANICA_1_2004contraviolencia.pdf (Consultado: 31-1-2023)

Ministerio del Interior. (2022). VioGén cumple 15 años con más de 700.000 casos analizados y 5,4 millones de valoraciones de riesgo realizadas. [online] Disponible en: https://www.interior.gob.es/opencms/eu/detalle/articulo/VioGen-cumple-15-anos-con-mas-700.000-casos-analizados-y-54-millones-de-valoraciones-de-riesgo-realizadas/ (Consultado: 31-1-2023)

Moreno Yuste, L. (2015). Sistemas de seguridad en los centros penitenciarios. [PDF] Disponible en: http://derechop-cp62.wordpresstemporal.com/wp-content/uploads/2019/09/SISTEMAS-DE-SEGURIDAD-EN-LOS-CENTROS-PENITENCIARIOS.pdf (Consultado: 31-1-2023)

NL Times. (2021). Politicians, parents respond to Dutch Cabinet collapse. [online] Disponible en: https://nltimes.nl/2021/01/15/politicians-parents-respond-dutch-cabinet-collapse/ (Consultado: 31-1-2023)

OCDE. (2019). Fraud and corruption in European structural and investment funds. A spotlight on common schemes and preventive actions. [PDF] Disponible en: https://www.oecd.org/gov/ethics/prevention-fraud-corruption-european-funds.pdf (Consultado: 31-1-2023)

ONU-Mujeres. (2022). Hechos y cifras: Poner fin a la violencia contra las mujeres. [online] Disponible en: https://www.unwomen.org/es/what-we-do/ending-violence-against-women/facts-and-figures (Consultado: 31-1-2023)

Portela, M., Castillo, C., Tolan, S., Karimi Haghighi M., y Andrés Pueyo, A. (2022). A Comparative User Study of Human Predictions in Algorithm-Supported Recidivism Risk Assessment. [PDF] Disponible en: https://arxiv.org/pdf/2201.11080v2.pdf (Consultado: 31-1-2023)

Savary, P. (2009). Francis Evrard condamné à 30 ans dont 20 ans de sûreté. Reuters. [online] Disponible en: https://www.reuters.com/article/ofrtp-france-justice-evrard-verdict-urge-idFRPAE59T0ON20091030 (Consultado: 31-1-2023)

Selbst, A. y Powels, J. (2017). Meaningful information and the right to explanation. [PDF] Disponible en: https://academic.oup.com/idpl/article/7/4/233/4762325 (Consultado: 31-1-2023)

Silver, E. y Miller, L. L. (2002). A Cautionary Note on the Use of Actuarial Risk Assessment Tools for Social Control. Crime and Delinquency. [online] Disponible en: https://journals.sagepub.com/doi/abs/10.1177/0011128702048001006 (Consultado: 31-1-2023)

Smee, B. (2021). Queensland police to trial AI tool designed to predict and prevent domestic violence incidents. The Guardian. [online] Disponible en: https://www.theguardian.com/australia-news/2021/sep/14/queensland-police-to-trial-ai-tool-designed-to-predict-and-prevent-domestic-violence-incidents/ (Consultado: 31-1-2023)

Tiry, E. y Kim, K. (2021). Measuring Risk Assessment Tool Performance. [PDF] Disponible en: https://www.urban.org/sites/default/files/publication/103863/measuring-risk-assessment-tool-performance.pdf (Consultado: 31-1-2023)

Agradecimientos

Autora

Tanya Álvarez dirige la investigación de Digital Future Society Think Tank sobre brechas digitales y digitalización del sector público. Aboga por una perspectiva interdisciplinar del impacto de la tecnología en la sociedad. Es graduada en Historia del Arte por el Swarthmore College y tiene un máster en Gestión del Patrimonio Cultural por la Universidad de Barcelona.

Entrevistador

Pablo Jiménez Arandia es periodista. Trabaja como freelance investigando y escribiendo sobre el impacto social de la tecnología. También ha producido y dirigido varios proyectos periodísticos, entre ellos dos pódcast para Digital Future Society Think Tank. Durante su carrera ha cubierto asimismo información económica, sobre migraciones y política, entre otros temas.

Entrevistados (ordenados por casos)

Estos estudios de caso se basan en entrevistas a los siguientes expertos y usuarios:

BOSCO

- Eva Belmonte, periodista y codirectora de la Fundación Civio
- David Cabo, codirector de la Fundación Civio
- Sergio Carrasco, ingeniero informático y licenciado en Derecho
- Nerea Caballero, asistente social de Insercoop
- Mercedes, beneficiaria del bono social

RisCanvi

- Antonio Andrés Pueyo, científico y catedrático de la Universidad de Barcelona
- Griselda Barris, Departamento de Justicia de la Generalitat de Catalunya
- Manuel Portela, investigador de la Universidad Pompeu Fabra
- Núria Monfort, abogada de IACTA

VioGén

- Pilar Álvarez, periodista del diario *El País*
- Sonia Márquez, abogada de la Fundación Ana Bella
- Belén Méndez, superviviente de violencia de género de la Fundación Ana Bella
- Gemma Galdón, fundadora de la Fundación Eticas

SALER

- Alfons Puncel, funcionario público y exsubsecretario de la Conselleria de Transparencia de la Generalitat Valenciana
- Pedro Giménez, inspector de servicios de la Generalitat Valenciana
- Oscar Capdeferro, profesor lector de la Universidad de Barcelona, Departamento de Derecho Administrativo

Equipo de Digital Future Society Think Tank

Gracias a la siguiente compañera de Digital Future Society Think Tank por sus aportaciones y su apoyo en la elaboración de este informe:

- **Olivia Blanchard**, investigadora de Digital Future Society Think Tank

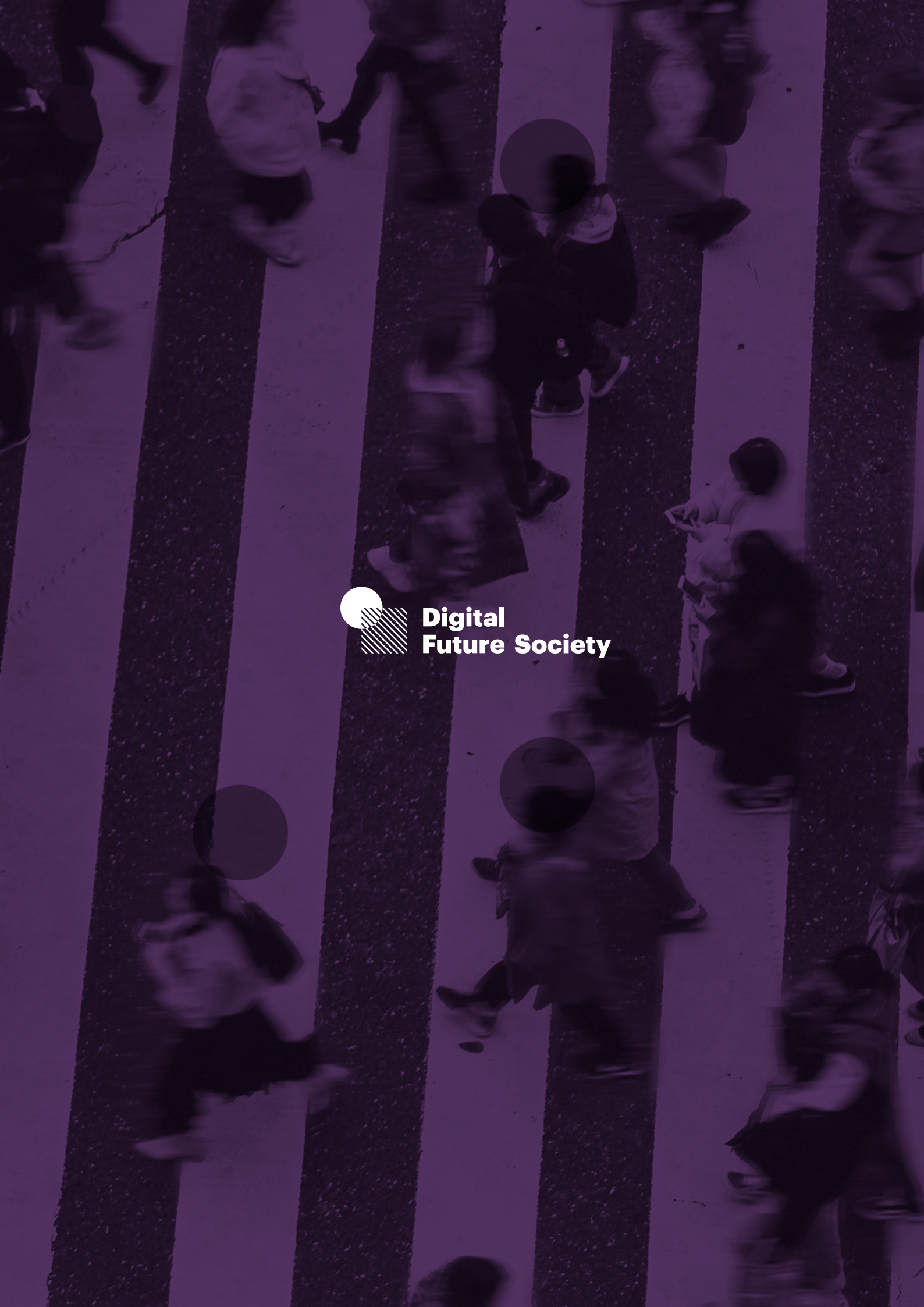
Citas

Este informe se debe citar de la siguiente manera:

- Digital Future Society. 2023. El uso de algoritmos en el sector público en España: cuatro estudios de caso sobre ADMS. Barcelona, España.

Datos de contacto

Si desea ponerse en contacto con el equipo de Digital Future Society Think Tank, envíe un correo electrónico a thinktank@digitalfuturesociety.com


Digital
Future Society