RESUMEN EJECUTIVO

Three predictive policing approaches in Spain: Viogén, RisCanvi and Veripol

Assessment from a human rights perspective

Lucía Martínez Garay (coord.)
Andrés Boix Palop, Álvaro Briz Redón
Fernando Flores Giménez, Andrea García Ortiz,
Mireia Molina Sánchez, Francisco Montes Suay,
Adrián Palma Ortigosa, Alfred Peris Manguillot,
Alba Soriano Arnanz

MINISTERIO DE CIENCIA E INNOVACIÓN
UNIÓN EUROPEA
FONDO EUROPEO DE DESARROLLO REGIONAL
AGENCIA ESTATAL DE INVESTIGACIÓN
Acció finançada per:
GENERALITAT VALENCIANA
Conselleria d'Innovació, Universitats, Ciència i Societat Digital
ALGORITHMICAL LAW
PROMETEO/2021/009

THREE PREDICTIVE POLICING APPROACHES IN SPAIN: VIOGÉN, RISCANVI AND VERIPOL (ASSESSMENT FROM A HUMAN RIGHTS PERSPECTIVE)

Resumen ejecutivo

El informe que se enlaza a continuación, "Three predictive policing approaches in Spain: Viogén, RisCanvi and Veripol (Assessment from a human rights perspective)", ha sido elaborado por un grupo de profesores de la Universitat de València y la Universitat Politècnica de València, especialistas en distintas ramas del Derecho (penal, administrativo y constitucional), la estadística y la matemática, bajo la coordinación de la Dra. Lucía Martínez Garay, Profesora de Derecho penal de la Universitat de València. El trabajo evalúa tres herramientas algorítmicas (con uso de Inteligencia Artificial, IA, en una de ellas), empleadas por la Administración española para realizar análisis predictivos en los ámbitos policiales y penitenciarios, desde la perspectiva de su respeto a los derechos fundamentales de los ciudadanos que son objeto de las evaluaciones para las que estas herramientas son empleadas. Los objetivos de este informe han sido: 1) analizar la capacidad predictiva de los tres sistemas y su eficacia para alcanzar el objetivo que pretenden; 2) analizar su grado de transparencia tanto para la ciudadanía en general como para las personas directamente afectadas por su utilización; 3) identificar posibles sesgos estadísticos, 4) analizar si respetan el vigente marco jurídico europeo y español, tanto en materia de protección de datos como respecto del empleo de este tipo de modelos, y, finalmente, 5) valorar si los costes para los derechos humanos exceden lo que sería legítimo, necesario y/o proporcionado en relación con el beneficio actual o probable en términos de prevención del delito. Aunque estos han sido los objetivos comunes para las tres herramientas, el análisis se ha tenido que adaptar a las características de cada una de ellas y, sobre todo, al tipo y a la cantidad de datos e información de que hemos podido disponer, que ha sido dispar en cada una de las tres. Asimismo, en el análisis se han combinado métodos cuantitativos y cualitativos, de nuevo de forma distinta para cada herramienta según la información a la que nos ha sido posible acceder.

GÉNESIS DEL INFORME Y ELECCIÓN DE LAS HERRAMIENTAS ESTUDIADAS

El origen de este informe es un encargo de una ONG de ámbito europeo e internacional que pretende realizar un estudio comparativo respecto de los riesgos que las herramientas de policía predictiva utilizadas en distintos estados europeos pueden plantear para los derechos fundamentales de los ciudadanos. El presente trabajo se refiere a la parte del análisis sobre las herramientas de uso predictivo empleadas en España, y ha sido realizado, en el marco de este encargo, a partir de un contrato de transferencia del conocimiento con la Universitat de València, bajo la dirección y coordinación de Lucía Martínez Garay. Los resultados del mismo son, además, y en parte, fruto de la tarea de investigación realizada por la mencionada profesora y el resto del equipo en varios proyectos de investigación («Algorithmical Law», PROMETEU/2021/009; PID2021-123441NB-I00, financiado por MCIN/

AEI/10.13039/501100011033/ y «ERDF A way of making Europe»; y «Digital economy regulation: equality guarantees provided by public powers and algorithmical tools», PID2019-108745GB-I00). En todo caso, los análisis, conclusiones y opiniones expresadas en este informe son únicamente los de sus autores y no representan los de las instituciones a las que pertenecen ni los de las que han financiado estas investigaciones, así como tampoco a la ONG que lo ha encargado.

Este informe no analiza todas las herramientas de esta índole que en estos momentos emplean las Administraciones españolas, sino únicamente tres de ellas, que seleccionó la ONG que encargó el informe a partir de un listado más amplio proporcionado por el equipo investigador. Fue también la ONG quien definió los aspectos que debían ser analizados en cada una. En concreto, se han estudiado dos herramientas algorítmicas utilizadas por la policía (VeriPol, empleada para la detección de posibles denuncias falsas de robo; y VioGén, utilizada para estimar los niveles de riesgo para las víctimas de violencia de género, a fin de decidir mejor las medidas de protección policial necesarias en cada caso) y una tercera que utiliza la administración autonómica catalana para la gestión de sus prisiones (RisCanvi, que estima tanto el riesgo de reincidencia como el de otras conductas peligrosas por parte de la población reclusa).

CONDICIONANTES Y LÍMITES DE LA REALIZACIÓN DEL INFORME

Este informe fue finalizado y entregado en noviembre de 2022, aunque su publicación por parte de la ONG que lo encargó, junto al resto de informes de otros países europeos, aún no ha tenido lugar. No obstante, en esta versión se han añadido, aunque de manera sucinta y fragmentaria, ciertos datos e información sobre VioGén y RisCanvi obtenidos hasta abril de 2023, así como algunas referencias a la auditoria que las propias autoridades autonómicas catalanas encargaron respecto de RisCanvi y que ha sido publicada en el primer trimestre de 2024. Estas fechas han condicionado también el marco normativo de referencia respecto del que se ha evaluado la compatibilidad de las herramientas con los derechos fundamentales: la regulación española y europea vigente en materia de empleo de herramientas automatizadas por los poderes públicos, legislación sobre modelos predictivos, y normativa sobre protección de datos de carácter personal, vigente en 2022. Por esta razón, el informe no evalúa el grado de cumplimiento de las herramientas con las obligaciones que el Reglamento Europeo en materia de Inteligencia Artificial (AI Act), aprobado por el Parlamento Europeo el 13 de marzo de 2024 y publicado por el Diario Oficial de la Unión Europea en junio de 2024, establece respecto del uso de sistemas algorítmicos por parte de los poderes públicos. Téngase en cuenta, a estos efectos, que la limitada definición de “Inteligencia Artificial” que el art. 3 de este Reglamento contiene en su versión final, según la cual esta requiere de un sistema diseñado para funcionar no sólo con diversos niveles de autonomía sino que, además, sea capaz de mostrar capacidad de adaptación tras su despliegue, es decir una capacidad de evolución/aprendizaje mínimamente autónomo, hace que de las tres herramientas analizadas sólo VeriPol pueda entenderse que entra en la actualidad en sentido estricto en esta definición y debería, por tanto, cumplir con todas las previsiones del reglamento cuando entre en vigor para los usos de

IA por parte de los poderes públicos (lo que no está previsto hasta dentro de unos años por el propio reglamento, en todo caso, lo que hace que este análisis sea más importante en cuanto a pautas y directrices a aplicar que de cumplimiento normativo estricto). Por lo demás, esto podría cambiar en un futuro, dado que hay, en el caso de VioGén, evoluciones en proceso dirigidas a incorporar moderadamente sistemas de IA para la mejora evolutiva de la herramienta a fin de incrementar su capacidad predictiva, y de que respecto de RisCanvi se han propuesto también modificaciones en este sentido (como resultado de la auditoría externa encargada por la Generalitat). En todo caso, la mayor parte de los juicios de valor sobre adaptación de las herramientas analizadas al marco jurídico y al cumplimiento de ciertas exigencias de transparencia y de auditabilidad, se corresponden con el grueso de las exigencias que esta norma jurídica de Derecho europeo ha incorporado para los supuestos de utilización de IA por parte de los poderes públicos. Pues entendemos que, si bien no como marco normativo cuyo cumplimiento sea ya directamente exigible a los poderes públicos, sí constituyen no obstante guías y directrices que, como mínimo, han de ser tenidas en consideración cuando emplean estas herramientas.

Por último, queremos señalar que la información sobre la que se ha elaborado el informe es en parte de acceso público, y en parte nos ha sido proporcionada por las instituciones que gestionan cada una de las herramientas analizadas, o bien por personas que las utilizan o conocen por su actividad profesional. Respecto de esta segunda clase de información, sólo hemos reproducido aquí los datos o extremos que no estaban expresamente cubiertos por un acuerdo de confidencialidad. También hemos mantenido en todo caso el anonimato de todas las fuentes empleadas cuando así lo han requerido y cuando su identificación personal no era relevante a los efectos del informe realizado.

CONCLUSIONES GENERALES APLICABLES A LAS TRES HERRAMIENTAS ANALIZADAS

Las principales conclusiones que se extraen del informe, y que son predicables de las tres herramientas analizadas, son las siguientes.

1. En general, todas ellas cumplen con el marco normativo vigente en España en materia de empleo de herramientas algorítmicas y sistemas automatizados para el apoyo a la acción administrativa, lo que en gran parte tiene que ver no con un diseño de las mismas particularmente riguroso y exigente en cuanto a sus garantías, sino con la muy limitada ambición del marco jurídico español vigente en esta materia. Igualmente, no se han detectado incumplimientos relevantes respecto de las obligaciones de Derecho europeo y nacional en materia de protección de datos de carácter personal, lo que también es consecuencia de que resulta relativamente fácil cumplirlas en estos casos porque el marco no es particularmente exigente con los poderes públicos (que han de cumplir con las exigencias técnicas pero ni siquiera siempre con la necesidad de consentimiento, que puede ser sustituido por una habilitación legal). En todo caso, en el informe se proponen mejoras en todos los casos en esta materia, que tienen que ver con una mejor auditoría y evaluación de los riesgos, que no parece haber sido realizada con la intensidad y profundidad exigida por el marco jurídico vigente, tanto en uno

como en otro ámbito, cuando estamos ante sistemas que pueden tener una afección notable a los derechos de los ciudadanos. Asimismo, entendemos que hay que garantizar mejor el derecho de acceso de los ciudadanos a sus datos personales empleados por el sistema.

2. El principal punto negro que comparten las tres herramientas, si bien en diversos grados, es la falta de transparencia, información y suministro de datos completos sobre su funcionamiento, que facilite una auditoría y control externos capaces de detectar eventuales sesgos discriminatorios o riesgos para los derechos de los ciudadanos, de una parte; y que permita, de otra, evaluar mejor la capacidad predictiva y su eficacia en la prevención del delito, sin tener que confiar para ello únicamente en los informes y evaluaciones de las propias administraciones públicas que las emplean o han diseñado. En relación con este punto es importante, a nuestro juicio, diferenciar dos cuestiones. Una cosa es que el diseño, la implementación o incluso el control del funcionamiento de estos sistemas haya sido riguroso, realizado por especialistas, muchas veces en saludable colaboración entre las Administraciones y las Universidades; esto es sin duda una primera garantía de que las herramientas están construidas con rigor y calidad desde el punto de vista técnico. Pero otra cosa, distinta, es que a pesar de ello siguen siendo necesarios tanto la transparencia de cara a la ciudadanía, como el control externo por parte de terceros no implicados en el diseño ni el funcionamiento diario de los sistemas. Lo primero, porque tanto el diseño estas herramientas como el establecimiento de las pautas sobre cómo usarlas, no sólo implican sólo decisiones técnicas, sino también decisiones sobre qué fines deben ser priorizados, y qué costes se considera aceptable asumir para alcanzar aquellos fines. Decisiones que son por tanto en última instancia de naturaleza política, y sobre las cuales los ciudadanos tienen derecho a estar informados, por razones de confianza de la ciudadanía en las herramientas y en los poderes públicos que las emplean, y para poder controlar la acción de los mismos. Especialmente en el caso de herramientas que, como las tres analizadas, se usan por los poderes públicos como apoyo para tomar decisiones que pueden redundar en restricciones de derechos fundamentales. Y lo segundo, porque también un principio básico del método científico es la publicidad de los métodos, de los datos y de los resultados, para que puedan ser confirmados y refutados por otros investigadores, lo que redunda en la detección y corrección de errores, la discusión y el debate bien informados y, en último término, la formulación de críticas y propuestas de mejora y el avance del conocimiento. Si una de las razones que se alegan para utilizar algoritmos en apoyo de la toma de decisiones es que la cuantificación y matematización de la información produce decisiones más objetivas y eficaces, y mejores porque están basadas en la evidencia científica, debe ser posible que dicha mayor objetividad y eficacia sean contrastables científicamente, lo que requiere compartir los datos, los códigos y los resultados.

   A este respecto, y de acuerdo con la información disponible, las tres herramientas parecen a priori construidas y diseñadas de manera seria y rigurosa. Sin embargo, todas presentan problemas en relación con la transparencia hacia los ciudadanos y hacia la comunidad de expertos, porque faltan datos esenciales para poder realizar una

auditoría externa completa totalmente fiable: no se encuentran disponibles todos los dataset, no se conoce el diseño del algoritmo ni el peso relativo de cada uno de los factores predictivos y, en el caso de VeriPol, es incluso complicado saber si se usa de manera generalizada y consistente por parte de las autoridades. Dicho esto, es necesario no obstante singularizar el caso de RisCanvi, que merece un juicio aparte y mucho más favorable, en la medida en que la Generalitat de Catalunya está haciendo esfuerzos muy apreciables en los últimos tiempos para suministrar en abierto mucha más información sobre la herramienta y su funcionamiento, como destacamos en el apartado correspondiente.

3. Además, la lógica del Reglamento Europeo de Inteligencia Artificial, al que ya hemos hecho referencia, obligaría a una mayor transparencia para el caso de que estas herramientas incorporaran en el futuro IA con capacidad de aprendizaje y adaptativa, y ello en dos sentidos. Por un lado, en cuanto a las obligaciones de una mejor y más completa explicación de la lógica de funcionamiento de las herramientas en abierto con carácter general. Y por otro lado, y con mucho más detalle, respecto de la información que será obligatorio suministrar a los controladores públicos y autoridades de verificación privadas, para que las administraciones públicas en el ámbito de la Unión Europea puedan emplear sistemas de esta naturaleza.

4. En cuanto a la capacidad predictiva y a la eficacia para alcanzar el objetivo propuesto de reducir la comisión de ciertos delitos o la prevalencia de determinadas conductas, es difícil pronunciarse de manera general porque la cantidad y la calidad de la información disponible sobre cada herramienta analizada es muy dispar, por lo que remitimos al análisis efectuado en el informe respecto de cada una en concreto. Lo único que destacaríamos aquí como consideración general es que las alegaciones de los responsables del despliegue de estos sistemas, en el sentido de que han mejorado los resultados previos a su introducción, parecen consistentes (con la excepción de VeriPol, donde es imposible cualquier verificación de estos extremos por la falta de información disponible), aunque tienen limitaciones importantes: en algunos casos la información y los datos sobre los que se sustentan estas afirmaciones son muy escasos, y en general es difícil hacer una comparación con lo que ocurriría en ausencia de estos sistemas. Pero, en todo caso, echamos en falta una reflexión más profunda que compare lo que se gana en términos de prevención del delito con el coste en cuanto a restricción de derechos fundamentales, para poder realizar un juicio sobre la proporcionalidad del empleo de estos sistemas. Juicio que además, en nuestra opinión, debería ser diferenciado según la decisión para la que se utilice el algoritmo: un determinado nivel de acierto en las estimaciones de riesgo puede ser suficiente para tomar decisiones beneficiosas para el condenado o acusado, pero quizá no para tomar decisiones que restrinjan sus derechos; o un determinado nivel de acierto puede bastar para adoptar medidas a nivel policial en favor de la víctima, y sin embargo sería discutible que bastara para imponer medidas cautelares al acusado. Consideramos que en este aspecto queda mucho por hacer, y que para que este necesario debate sea siquiera posible es imprescindible un mejor y más completo acceso a la información sobre los algoritmos, su funcionamiento y sus resultados, como hemos señalado supra.

5. En cuanto a posibles sesgos discriminatorios o impacto dispar de las estimaciones de riesgo sobre grupos diferentes de población, la información publicada por los responsables de estas herramientas es mucho más escasa aún que la disponible sobre capacidad predictiva, y también lo son los datos a partir de los cuales poder hacer una evaluación externa (si bien de nuevo aquí hay importantes diferencias entre las tres herramientas analizadas, siendo RisCanvi - otra vez - la que ofrece más y mejor información, y Veripol la más opaca). Nosotros hemos detectado algunos problemas en este sentido tanto en VioGen como en RisCanvi (y los intuimos en Veripol, pero aquí la ausencia de datos impide formular nada más que meras intuiciones), si bien las conclusiones que hemos alcanzado han de ser tomadas con mucha cautela, porque están elaboradas sobre una base muy endeble, como hemos tratado de subrayar a lo largo del informe. Ahora bien, precisamente esto es lo que a nuestro juicio resulta más preocupante: que a día de hoy (y con la importante excepción parcial de RisCanvi, como destacamos en el apartado correspondiente) resulte imposible estudiar algo tan relevante como los eventuales sesgos o efectos discriminatorios que estas herramientas pueden estar produciendo – o no – habida cuenta de la ausencia de datos de acceso público con los que trabajar. Que una cuestión de esta relevancia no pueda ser investigada por personas externas a quienes son responsables de la gestión de estas herramientas (y que, por otra parte, ni siquiera ha sido casi estudiada hasta ahora por estos responsables) nos parece un problema grave.

PRINCIPALES CONCLUSIONES RESPECTO DE VIOGÉN

Por lo que se refiere a VioGén, las principales conclusiones del estudio son la ya referida necesidad de una mayor información sobre su funcionamiento. Si bien hay varios trabajos sobre distintos aspectos del funcionamiento de la herramienta publicados en revistas académicas de prestigio, eso no permite el tipo de control externo necesario al que nos hemos referido supra. En particular, con VioGen podría seguirse un modelo de facilitación de datos anonimizados en abierto similar al que se ofrece, sin problema alguno, con RisCanvi (véase infra). O podrían incluirse en el portal estadístico de la web de la Delegación del Gobierno contra la Violencia de Género muchos más datos de los que ahora tiene, que son escasos y aportan poca información. Ello impide una evaluación rigurosa y fiable del funcionamiento de la herramienta desde fuera, tanto en cuanto a su eficacia y fiabilidad como respecto de la identificación de posibles problemas y riesgos.

Con todo, de la información existente se desprende que las afirmaciones de sus responsables sobre el buen funcionamiento de la herramienta en cuanto a su capacidad para reducir la revictimización de las mujeres que denuncian haber sufrido violencia de género parecen consistentes con los datos disponibles en abierto, y con la ampliación de los mismos que nos ha sido proporcionada. En particular, creemos necesario ser prudentes en relación con las críticas que ha recibido VioGen por no ser capaz de detectar mejor los casos en que la mujer está en riesgo de homicidio a manos de su pareja o expareja. En nuestra opinión, el Ministerio ha realizado un indiscutible esfuerzo por mejorar la predicción de esos casos, que por otra parte

son muy difíciles de detectar dada su bajísima prevalencia, e implican siempre una tasa muy elevada de falsos positivos. Sin perjuicio de las mejoras que puedan seguir haciéndose en este ámbito (y que no necesariamente tienen que pasar siempre por la actuación policial o judicial) nos parece necesario ser conscientes de los límites a que la predicción de este tipo de eventos está sometida, y reconocer que ningún sistema de estimación de riesgos puede aspirar razonablemente a reducir estas tragedias a cero.

La capacidad predictiva de VioGen es buena por lo que hace a la detección del riesgo bajo (en el sentido de que en la inmensa mayoría de los casos de riesgo bajo no hay agresiones posteriores), pero más débil por lo que se refiere a los casos de riesgo alto, pues muchos casos en los que se afirma riesgo alto o extremo no van seguidos tampoco de agresiones ulteriores. Sin perjuicio de reconocer que la propia implementación de medidas de protección contribuye seguramente en parte a este efecto, hay que ser conscientes de que el sistema, como reconocen sus propios responsables, asume un nivel elevado de falsos positivos para poder maximizar la sensibilidad (esto es, no dejar sin protección a ninguna mujer que realmente pueda estar en riesgo). Y esta opción, que a nivel policial no es problemática desde la perspectiva de los derechos fundamentales porque las medidas de protección de las mujeres inciden muy poco en los derechos de los varones denunciados, resultaría sin embargo mucho más cuestionable si se utilizaran los niveles de riesgo de manera automática en el ámbito judicial para fundamentar la imposición de medidas cautelares, especialmente las más restrictivas de derechos. Esta utilización generalizada no está ocurriendo a día de hoy, al menos hasta donde hemos podido averiguar, pero es algo frente a lo que conviene estar alerta porque, en primer lugar, ya en la actualidad es posible tener en cuenta el resultado de VioGén para tomar estas decisiones, aunque no se perciba como determinante; y en segundo lugar, porque es probable que este uso pase a ser mayor en el futuro. En tal caso, el juicio emitido en este sentido debiera ser, como es obvio, notablemente revisado.

Por lo que hace a posibles sesgos o efectos discriminatorios de VioGén, ya hemos dicho que la muy limitada información disponible dificulta mucho poder analizar esta cuestión con un mínimo de rigor. Nosotros hemos detectado indicios que apuntan a una protección menor, o menos eficiente, de las mujeres nacidas fuera de España tanto frente a nuevas agresiones de género, como frente a homicidios. Señalamos en el estudio la metodología empleada para analizar los datos disponibles y las razones por las que llegamos a esta conclusión, que ciertamente se basa en información incompleta, pues incompleta es la información disponible al respecto, pero que consideramos que, en ausencia de más datos, y más completos, debería llevar al menos a una revisión del funcionamiento de la herramienta, tanto para la estimación del riesgo de violencia en general, como para la estimación del riesgo de homicidio con la nueva escala H.

En cuanto al cumplimiento normativo, se echan en falta evaluaciones más completas de riesgos, tanto en cuanto a los datos de carácter personal empleados y los riesgos que de su uso puedan derivarse como, sobre todo, en materia de las evaluaciones sobre riesgos para derechos fundamentales y de prevención de sesgos que un sistema de este tipo debería incorporar. De hecho en el modelo de cumplimiento del Reglamento Europeo en materia de Inteligencia Artificial estas evaluaciones pasarán a ser obligatorias para sistemas de esta índole

desde el momento en que incorporen IA con capacidad de evolución o aprendizaje (lo que la versión actual de VioGén parece que no hace, pero que en cuanto pase a ser empleada una que sí lo haga deberá adaptarse a esta exigencia). Por último, también son susceptibles de una mejora importante los mecanismos que garanticen el acceso efectivo de los ciudadanos a sus evaluaciones y a todos sus datos de carácter personal almacenados por el sistema, algo que en la actualidad no se garantiza suficientemente.

PRINCIPALES CONCLUSIONES RESPECTO DE RISCANVI

RisCanvi es la herramienta más transparente de las analizadas. Si bien el diseño de los algoritmos se mantiene secreto (con la excepción del que estima el riesgo de reincidencia violenta, que hemos podido conocer gracias a la colaboración de colegas de la Universitat Pompeu Fabra) la Generalitat de Cataluña ha ido poniendo en práctica, de manera creciente con los años, una política de open data con publicación de datos anonimizados en formato excel en su página web, que es reseñable en un sentido muy positivo, que creemos que debería servir de modelo a otras Administraciones, y que permite analizar el funcionamiento de los algoritmos a cualquier investigador interesado sin comprometer en modo alguno la privacidad de las personas afectadas por las estimaciones de riesgo. Además, la herramienta ha sido auditada en cuanto a su funcionamiento recientemente, y los resultados de esta evaluación han sido también publicados. Desde este punto de vista, si en un futuro la herramienta incorporara para la mejora de sus estimaciones de riesgo sistemas de IA en el sentido definido por el Reglamento europeo, probablemente podrá cumplir las exigencias de transparencia y auditabilidad del sistema con bastante facilidad.

Queremos subrayar que de los cinco tipos de riesgo que estima RisCanvi, en este informe sólo hemos analizado el de reincidencia violenta, por lo que todas nuestras consideraciones se limitan exclusivamente a este aspecto.

La capacidad predictiva de RisCanvi es bastante buena en cuanto a la detección de las personas con riesgo bajo, pero mucho menor en cuanto a la detección de las personas con riesgo alto. Las estimaciones de riesgo incluyen un número elevado de falsos positivos y en esto coinciden tanto las investigaciones llevadas a cabo desde organismos dependientes de la Generalitat, como los cálculos que nosotros hemos podido realizar a partir de los datos disponibles, como también los resultados de la autoría externa. Creemos que los responsables de la herramienta son conscientes de esta realidad y están extrayendo de ella la conclusión, razonable desde nuestro punto de vista, de que las evaluaciones de riesgo deberían conducir a una suavización de los regímenes de vida y a mayores posibilidades de acceso a la libertad para las personas de riesgo bajo, pero no llevar automáticamente a un endurecimiento del régimen ni de las condiciones de cumplimiento para aquellos clasificados como de riesgo alto. Sería deseable, no obstante, que existieran pautas claras de actuación a este respecto, y también que este conocimiento se extendiera al ámbito judicial, donde se revisan las decisiones de clasificación de las Juntas de Tratamiento, y donde la Fiscalía tiende a dar una importancia muy elevada a

las clasificaciones de riesgo alto o medio, sin tener en cuenta las importantes limitaciones que esas estimaciones tienen.

Algunos de los factores que Riscanvi tiene en cuenta para evaluar el riesgo afectan muy directamente a la vida privada e intimidad de las personas y a su libre desarrollo de la personalidad o a su entorno de socialización. Puesto que no son públicos los algoritmos (con la excepción antes mencionada) no se puede saber cuál es la capacidad predictiva de esta información tan sensible. Y sin embargo puede resultar cuestionable que aspectos sobre los que la persona no tiene control (antecedentes delictivos de sus familiares, por ejemplo) o forman parte de su derecho al libre desarrollo de la personalidad (número y tipo de actividades sexuales preferidas, por ejemplo) puedan constituir una base legítima para fundamentar medidas restrictivas de derechos (por ejemplo, la denegación del acceso a un tercer grado o a la libertad condicional). Sería necesario comparar el rendimiento de estos ítems (es decir, cuánto de pierde en cuanto a capacidad predictiva si se eliminan del algoritmo) con su coste en términos de afectación a los derechos (a la intimidad, a la igualdad, al respecto a la vida privada y familiar, etc.) para tomar una decisión sobre si resulta proporcionado emplearlos en las predicciones. Además, en algunos de los ítems puede incluso cuestionarse que sea pertinente tenerlos en cuenta incluso en el supuesto de que tengan valor predictivo (por ejemplo, la presencia de discapacidad). Creemos que sería necesaria una auditoría profunda respecto del riesgo que comportan estos elementos para los derechos fundamentales de los reclusos, que no nos consta que se haya llevado a cabo, así como una transparencia mucho mayor sobre las evidencias empíricas que, en su caso, podrían avalar que se entiendan estos factores como estadísticamente significativos.

Por último, nuestro análisis estadístico ha revelado cierto trato extraordinariamente severo para algunos colectivos de reclusos, que no parece justificado y que podría retroalimentar sesgos en la herramienta en perjuicio de estos colectivos, esencialmente las personas con problemas de salud mental o de adicciones, o personas con una situación socioeconómica muy desfavorable. No así, por el contrario, y creemos que esto es una virtud de RisCanvi que debe ser destacada, en cuanto a un trato más severo a los extranjeros, que no se da según los datos disponibles. Entendemos que el funcionamiento de la herramienta debería al menos revisarse a partir de los datos estadísticos que aportamos, y valorarse alguna vía para solucionar estos problemas.

PRINCIPALES CONCLUSIONES RESPECTO DE VERIPOL

VeriPol es la herramienta de la que menos información se dispone, lo que agrava el problema de falta de transparencia y las dificultades, cuando no imposibilidad, para realizar la tarea de control externo que hemos pretendido con este informe. Agradecemos a la Policía Nacional la información y los datos que nos proporcionó para al menos podernos hacer alguna idea sobre esta herramienta y el tipo de uso que se le da en la práctica, pero aun así se trata de una situación muy deficiente en cuanto al cumplimiento de los estándares mínimos de transparencia que consideramos requeridos para el empleo de herramientas de este tipo, y

aunque en el informe hemos analizado toda la información de la que hemos podido disponer, muchas conclusiones no pasan de ser conjeturas, a falta de datos contrastables con los que poder trabajar.

En cuanto a los problemas que pueden derivarse de su afección a derechos fundamentales o en cuanto a la posible generación de sesgos, esta misma falta de información impide una valoración externa sobre el funcionamiento práctico del sistema. El informe ha tratado, al menos, de establecer un marco teórico respecto de las cautelas mínimas exigibles para evitar que modelos como éste, basados en IA y en aprendizaje a partir de sistemas de procesamiento de lenguaje natural, incurran en sesgos para las funciones que se le presuponen (detección de denuncias falsas por parte de los ciudadanos y ciudadanas), que, por ejemplo, puedan afectar más a la parte de la población que maneja peor el idioma por razones que pueden ser de muy diverso tipo (nivel socioeconómico, origen, nivel de estudios, e incluso la mera diferencia diatópica dialectal en casos en que pueda emplearse una variedad diferente a la que ha sido mayoritariamente empleada para entrenar el modelo). Asimismo, y como para el resto de herramientas, trata de sugerir mejoras para una mejor evaluación de los posibles riesgos para los derechos fundamentales que puedan derivarse de su utilización, tantos más cuanto más común sea ésta.