

CODDOU MC MANUS, Alberto; RUIZ-ESQUIDE ENRÍQUEZ, Isabel; BERGERET SÁEZ, Dominique; LOYOLA OLMOS, Camila; SÁNCHEZ ANDRADE, Valentina: “IA y violencia contra las mujeres: la automatización de la evaluación del riesgo”.
Polít. Crim. Vol. 20 N° 40 (Diciembre 2025), Art.14, pp. 346- 383
<https://politicrim.com/wp-content/uploads/2025/11/Vol20N40A14.pdf>

IA y violencia contra las mujeres: la automatización de la evaluación del riesgo

AI and gender-based violence: the automation of risk assessment

Alberto Coddou Mc Manus

Doctor en Derecho

Profesor asociado Escuela de Gobierno de la Pontificia Universidad Católica de Chile

acoddou@uc.cl

<https://orcid.org/0000-0003-2041-2304>

Isabel Ruiz-Esquide Enríquez

Doctora en Derecho

Docente de la Facultad de Derecho y Ciencias Sociales de la Universidad San Sebastián

isaruize@yahoo.es

<https://orcid.org/0000-0003-3957-6771>

Dominique Bergeret Sáez

Estudiante de Derecho, Universidad Austral de Chile

dominique.bergeret@alumnos.uach.cl

<https://orcid.org/0009-0005-2920-356X>

Camila Loyola Olmos

Estudiante de Derecho, Universidad Austral de Chile

camila.loyola@alumnos.uach.cl

<https://orcid.org/0009-0000-6469-0408>

Valentina Sánchez Andrade

Estudiante de Derecho, Universidad Austral de Chile

valentina.sanchez01@alumnos.uach.cl

<https://orcid.org/0009-0007-2430-7778>

Fecha de recepción: 21/08/2024

Fecha de aprobación: 16/05/2025

Resumen

El presente artículo tiene por objeto principal analizar críticamente, y desde una perspectiva de derechos humanos aplicables a las nuevas tecnologías, la implementación piloto de un algoritmo de sugerencia de riesgo (ASR) en el Ministerio Público chileno. Para ello, describiremos los factores que explican su desarrollo, la experiencia previa de la Pauta Unificada de Evaluación Inicial del Riesgo (PUIR), la experiencia y las lecciones que pueden extraerse del VIOGEN español, y realizaremos una explicación detallada del diseño y modo de funcionamiento del ASR. Incorporamos una metodología cualitativa para analizar cómo los funcionarios del Ministerio Público interactúan con esta herramienta, en esta fase de implementación. Dado que el ordenamiento jurídico chileno está en transformación, con nuevas regulaciones sobre protección de datos y el posible uso de sistemas de IA, consideramos crucial contribuir a conciliar dos aspectos fundamentales: por un lado, mejorar la evaluación del riesgo de violencia contra las mujeres y hacer más eficientes las medidas de protección; y, por otro, proteger los derechos fundamentales de víctimas e imputados en un contexto de creciente intermediación algorítmica entre Estado y ciudadanos.

Abstract

The main objective of this article is to critically analyze, from a human rights perspective applicable to new technologies, a risk-assessment algorithm (ASR) currently being piloted by the Chilean Public Prosecutor's Office. To achieve this, we will describe the factors that explain its development, the prior experience with the Unified Initial Risk Assessment (PUIR), the experience and lessons that can be drawn from the Spanish VIOGEN system, and provide a detailed explanation of the design and functioning of the ASR. We incorporate a qualitative methodology to analyze how Public Prosecutor's Office officials interact with this tool, still in its pilot stage. Given that the Chilean legal framework is undergoing changes, with new regulations on data protection and the potential regulation of AI systems, we consider it essential to contribute to the reconciliation of two key aspects: on the one hand, improving risk assessments for violence against women and making protection measures more efficient; and, on the other, protecting the fundamental rights of victims and defendants in an era marked by increasing algorithmic mediation in the vertical relationships between the State and its citizens.

Palabras clave: Inteligencia Artificial, violencia de género, algoritmos públicos

Keywords: Artificial Intelligence, gender-based violence, public algorithms

Introducción¹

La violencia contra las mujeres ha sido descrita como una “epidemia” de carácter “endémica”, “omnipresente” y “devastadora”.² A nivel global, una de cada tres mujeres informa haber experimentado violencia física o sexual en manos de sus parejas.³ En Chile, el Presidente de la República describió la violencia contra las mujeres como un problema de “seguridad pública”.⁴ Más allá de los esfuerzos desplegados por movimientos de mujeres, organizaciones públicas,

¹ Este trabajo fue desarrollado en el marco de los proyectos FONDECYT de Iniciación N° 11220370, FONDECYT Regular N° 230895 y FONDECYT Regular N° 1240227.

² Organización Mundial de la Salud (2021), *passim*.

³ BUZAWA y BUZAWA (2017), *passim*.

⁴ Gobierno de Chile (2023), *passim*.

privadas, nacionales e internacionales, que han establecido tratados internacionales, convenios, acuerdos, políticas públicas, entre otras medidas, la violencia contra las mujeres no ha sido erradicada ni significativamente reducida.⁵ En 2022, la Subsecretaría de Prevención del Delito dio cuenta de 123.265 denuncias por situaciones enmarcadas en violencia intrafamiliar (VIF). De ese total, 90.920 denuncias fueron realizadas por mujeres.⁶ Asimismo, la Fiscalía Nacional estimó que el 42% de las mujeres víctimas de homicidio son asesinadas en contexto de VIF, dejando en evidencia que la violencia de género y, en específico, la violencia en contra de las mujeres (VCM) es un problema que afecta sus derechos fundamentales, principalmente el derecho a una vida libre de violencia.⁷ Durante el año 2020, el Ministerio de la Mujer y la Equidad de Género registró 43 femicidios consumados y 151 femicidios frustrados en Chile.⁸ Respecto de los femicidios consumados, el 44,2% contaba con una denuncia anterior por violencia intrafamiliar, mientras que, en el 9,3% de los casos, la víctima tenía una medida cautelar vigente a su favor.⁹ En Chile, cuando las mujeres denuncian a sus parejas por haber sufrido algún tipo de violencia, ello no es garantía de que el Estado las protegerá contra nuevas agresiones, incluyendo lesiones graves o la propia muerte.

Para abordar este problema, el ordenamiento jurídico chileno ha reaccionado a través de diversas vías, gradualmente mejorando la capacidad que tiene el Estado de proteger efectivamente a las víctimas y castigar a los agresores. Con la dictación de la Ley N° 20.066 (2005), que establece la ley de violencia intrafamiliar, se introdujeron medidas específicas de sanción que modifican el enfoque “terapéutico” anterior hacia un enfoque “represivo” con el agresor y de “protección” a las víctimas.¹⁰ En el artículo 5 de este cuerpo legal se define la VIF de un modo que enfrenta la violencia contra las mujeres “no en su condición de tales, sino en su rol dentro de las relaciones afectuosas”.¹¹ Más allá de las críticas que ha recibido esta definición, los hechos constitutivos de VIF pueden ser sancionados a través de un procedimiento especial ante los tribunales de familia o, en caso de que constituyan un delito previsto en la propia Ley N° 20.066 (por ejemplo, el delito de maltrato habitual) o en otro cuerpo legal, ser conocidos en el marco de un proceso penal.¹² Adicionalmente, la ley 20.066 establece una serie de normas especiales sobre el proceso penal a ser aplicado en caso de estar ante VIF constitutiva de delito, como la obligación de considerar la conducta anterior del agresor respecto a las anotaciones del registro sobre sanciones y medidas accesorias (artículo 12), la facultad de los jueces de garantía y del tribunal oral en lo penal de decretar medidas cautelares en cualquier momento del juicio (artículo 15), sumado a la prohibición de celebrar acuerdos reparatorios (artículo 19). En cuanto a las medidas cautelares, posibles de adoptar

⁵ Véase la Encuesta de Violencia contra la Mujer en el Ámbito de Violencia Intrafamiliar y en Otros Espacios, que se ha realizado en tres ocasiones (2012, 2017 y 2020) y que es gestionada por el Centro de Estudios y Análisis del Delito.

⁶ Véase las “Estadísticas Delictuales año 2022”, que recopila el Centro de Estudios y Análisis del Delito en función de los casos policiales, detenciones flagrantes y denuncias realizadas ante Carabineros de Chile o la Policía de Investigaciones de Chile.

⁷ Fiscalía Nacional (2022), pp. 1-25.

⁸ Circuito Intersectorial de Femicidio (2021), p. 10.

⁹ Circuito Intersectorial de Femicidio (2021), p. 20.

¹⁰ CASAS y VARGAS (2011), pp. 133-151.

¹¹ VILLEGAS (2017), p. 17.

¹² CARRASCO (2018), pp. 209-231.

durante la vigencia del procedimiento, el artículo 15 señala que son todas aquellas “necesarias para proteger a la víctima de manera eficaz y oportuna”, tales como las que se establecen en ley N°19.968, que crea los Tribunales de Familia,¹³ o en la recientemente aprobada ley que estatuye medidas para prevenir, sancionar y erradicar la violencia en contra de las mujeres, en razón de su género.¹⁴ Además, la ley 21.378, que introduce una medida cautelar especial en favor de las víctimas de VIF, faculta al tribunal, en casos calificados, a controlar su cumplimiento por medio del monitoreo telemático. Para decretar medidas de protección sobre la víctima, o cautelares en contra de los imputados, sobre todo en casos de riesgo inminente de sufrir maltrato constitutivo de VIF, el artículo 7 de la Ley 20.066 señala cuatro hipótesis que permiten presumir esta situación de riesgo con el sólo mérito de la denuncia o demanda.¹⁵

Para otorgar una efectiva protección a las víctimas de violencia de género, específicamente ante la violencia que tiene lugar en el ámbito doméstico o en relaciones de pareja, los Estados han ido progresivamente sofisticando sus herramientas de evaluación del riesgo, sobre todo para evitar la revictimización y o reportar un incidente ante las autoridades. Tradicionalmente, los procesos de evaluación del riesgo se hacían únicamente a través del juicio profesional de un experto, que evaluaba la vulnerabilidad de la víctima y/o las características del agresor para determinar la procedencia de una medida de protección. Sin embargo, a partir del manejo de una gran cantidad de casos que presentan patrones comunes, los sistemas de evaluación del riesgo han ido progresivamente utilizando las grandes cantidades de información que el propio sistema provee. A partir de esta información, diversos estudios criminológicos han ido observando y evaluando cuáles son los indicadores que permitirían predecir de manera más eficaz los casos de revictimización o de ocurrencia de nuevos hechos de violencia. Por ejemplo, algunos autores identifican las covariables que influyen significativamente en la predicción de la violencia de pareja, destacando el miedo al compañero como el factor más importante,¹⁶ seguido de las actitudes hacia la violencia, el historial de abuso y el consumo de alcohol y drogas, con el miedo siendo el predictor más fuerte de riesgo inminente y mostrando causalidad inversa (endogeneidad).¹⁷

¹³ De acuerdo con el artículo 92 de la Ley N.º 19.968, el juez de familia puede prohibir al agresor acercarse a la víctima y restringir su presencia en lugares frecuentados por ella, asegurar la entrega de los efectos personales de la víctima si decide no regresar al hogar común, fijar alimentos provisorios, establecer un régimen provisorio de cuidado personal de menores y la relación entre progenitores e hijos, prohibir la celebración de actos o contratos, prohibir el porte y tenencia de armas de fuego y otros objetos peligrosos, reservar la identidad del denunciante o implementar medidas de protección para adultos mayores o personas con incapacidad discapacidad.

¹⁴ Ley N.º 21.675, Estatuye medidas para prevenir, sancionar y erradicar la violencia en contra de las mujeres, en razón de su género, publicada en Diario Oficial el 14 de junio de 2024.

¹⁵ El inciso 2º de la norma citada establece cuatro hipótesis de riesgo inminente de sufrir maltrato intrafamiliar: intimidación previa, expresada mediante hostigamiento, acecho o invasión de espacios de la víctima; que el agresor presenta antecedentes como drogadicción, alcoholismo, denuncias o condenas por violencia intrafamiliar o delitos contra las personas; que el agresor se opone al fin de una relación afectiva con la víctima, empleando violencia física o psicológica; o que el agresor expulsa, relega o limita el desplazamiento de una persona mayor en su propia vivienda. Además, se enfatiza protección en situaciones vulnerables como embarazo, niñez o discapacidad.

¹⁶ AMUSA *et al.* (2022), *passim*.

¹⁷ NOVITZKY *et al.* (2023), p. 5.

A nivel comparado, la mayor parte de las herramientas de evaluación de riesgo utilizadas por las policías u otras autoridades se basan en cuestionarios breves que no requieren formación clínica y que son completados por los funcionarios pertinentes después de entrevistar principalmente a la víctima.¹⁸ A nivel global, los ejemplos son múltiples: desde el año 2007, España desarrolló un sistema centralizado a nivel nacional (VIOGEN) para la evaluación dinámica y continua de la materia, en la que están involucrados diversos organismos públicos;¹⁹ en Suecia, la policía nacional introdujo B-Safer en la década de 1990, que perfeccionaba un programa anterior (SARA);²⁰ siguiendo el modelo español, Portugal desarrolló su propio sistema, intentando corregir algunas de las deficiencias más evidentes del VIOGEN;²¹ por su parte, las fuerzas policiales en el Reino Unido han estado utilizando una herramienta estandarizada llamada DASH (*Domestic Abuse, Stalking and Honour Based Violence Risk Identification*), la que probablemente ha sido objeto del mayor número de evaluaciones y críticas.²²

En este escenario, la aplicación de sofisticadas tecnologías digitales, incluyendo herramientas de IA, se ha ido desarrollando sobre décadas de experiencia de las autoridades civiles y policiales en la gestión del riesgo de las mujeres víctimas de violencia. Si bien la “algoritmización de la violencia contra la mujer” es un fenómeno que venimos observando desde hace varias décadas,²³ la incorporación de herramientas de IA (como herramientas de *machine learning* o de *deep learning*) ha impactado significativamente en el modo en que las autoridades abordan la violencia contra la mujer. Si este fenómeno es descrito como una epidemia, y las presiones por disminuir y hacer más eficiente el gasto público en el tratamiento de este tipo de casos son cada vez mayores, entonces, se trata de un campo fértil para la introducción de sistemas automatizados de decisión.

A partir de esta revisión, nos planteamos las siguientes preguntas de investigación: ¿Qué explica que el Ministerio Público chileno decida diseñar e implementar un algoritmo para automatizar el proceso de evaluación de riesgo de las víctimas de violencia de género? ¿De qué manera interactúan los funcionarios públicos con esta herramienta durante su etapa de pilotaje? ¿Cuáles son los desafíos que plantea el ASR desde una perspectiva de derechos fundamentales, en particular en relación con la protección de datos personales y la no discriminación? Para responder estas preguntas, el presente artículo tiene por objeto principal analizar críticamente, y desde una perspectiva de derechos humanos aplicables a las nuevas tecnologías, la utilización del ASR en etapa de pilotaje. Para ello describiremos los factores que motivaron su desarrollo, la experiencia previa de la Pauta Unificada de Evaluación Inicial del Riesgo (PUIR), los aprendizajes que pueden extraerse del VIOGEN en España, así como también el diseño y modo de funcionamiento del ASR. En cuanto a la utilización de este algoritmo, incorporamos una metodología cualitativa que pretende explicar de mejor manera el modo en que los funcionarios del Ministerio Público han interactuado con esta herramienta. Nuestros hallazgos sugieren que,

¹⁸ TURNER *et al.* (2022), pp. 145-157.

¹⁹ LÓPEZ DE ZUBIRÍA DÍAZ (2024), pp. 77-94.

²⁰ SVALIN *et al.* (2018), pp. 71-79.

²¹ MACHADO, *et al.* (2021), pp. 215-238.

²² ALMOND, *et al.* (2017), pp. 58-68.

²³ LÓPEZ DE ZUBIRÍA DÍAZ (2024), p. 82.

aunque el ASR representa un avance técnico significativo que es de gran ayuda para los funcionarios encargados de su implementación, se requieren mecanismos efectivos de transparencia y explicabilidad, además de herramientas que permitan una adecuada evaluación de impacto. Por otra parte, si bien se constatan algunos riesgos relativos al sesgo de automatización y otras debilidades institucionales, el hecho de que la implementación del ASR descansa en funcionarios con expertise en la materia, permite minimizar algunos de los riesgos que se han observado a propósito del VIOGEN español.

Considerando que esta herramienta se encuentra en etapa de pilotaje, y de que el ordenamiento jurídico chileno está actualmente observando diversos cambios significativos en su estrategia de regulación de las nuevas tecnologías, principalmente por la nueva legislación en materia de protección de datos personales y por la eventual regulación del uso de los sistemas de IA, creemos necesario hacer una contribución que permita conciliar los dos bienes que están en juego: por un lado, utilizar las mejores herramientas disponibles para la evaluación del riesgo de violencia contra las mujeres, haciendo más eficientes las medidas de protección correspondientes; y, por otro, respetar y proteger los derechos fundamentales de las personas (víctimas e imputados) en una era marcada por la intermediación algorítmica de las relaciones verticales entre Estado y ciudadanos.

1.La evaluación del riesgo de las víctimas de VIF en el sistema procesal penal

En el caso chileno, y desde 2013, la División de Atención de Víctimas y Testigos de la Fiscalía Nacional (DAVT) evalúa el riesgo de las víctimas de VIF a través del uso de la Pauta de Evaluación de Riesgo en Violencia Intrafamiliar. Durante los años 2013 y 2014, este instrumento fue objeto de revisión y estudio por parte de la Universidad Cardenal Raúl Silva Henríquez.²⁴ Dicha casa de estudios habría realizado un análisis de confiabilidad y validez del instrumento, proponiendo ajustes que habrían sido realizados e implementados en el año 2015. Es interesante destacar que, pese a su importancia, no existe información respecto de los procesos de verificación de este instrumento, como tampoco ello se encuentra disponible en redes abiertas. El año 2016 se suscribió un Convenio entre el Ministerio Público, Carabineros de Chile, la Policía de Investigaciones y el Servicio Nacional de la Mujer y Equidad de Género. Todos ellos acordaron aplicar esta Pauta en el primer contacto que sostuvieran con las víctimas de VIF, esto es, mujeres mayores de 18 años con vínculos de pareja con su agresor. Se optó por la aplicación de la Pauta de Evaluación de Riesgo del Ministerio Público, por ser la única validada de manera técnica, convirtiéndose así en la Pauta Unificada Inicial de Riesgo (PUIR). La utilización de una metodología uniforme para los diversos operadores en la materia (policías y organismos estatales), ha sido destacada positivamente por el Banco mundial, aun cuando haya cuestionado que su aplicación no es uniforme.²⁵

²⁴ BRAVO y BARRA (2022), pp. 1-28.

²⁵ Banco Mundial (2021), pp. 1-85.

En términos concretos, la PUIR consiste en un cuestionario estructurado que contiene preguntas con puntaje jerarquizado y ponderado.²⁶ Su propósito es obtener, directamente de la víctima, información sobre la existencia de factores de riesgo asociados a la situación de violencia denunciada. El nivel cuantitativo de riesgo se calcula automáticamente, y permite calificar el riesgo como Vital/Alto, Medio o Bajo.²⁷ Con esa información, los operadores pertinentes deben adoptar las medidas y acciones de protección, según lo que se define en el Protocolo de Actuación, en concordancia con las instrucciones impartidas por el Fiscal Nacional.²⁸ Estas medidas pueden consistir en la necesidad de solicitar al tribunal competente la adopción de medidas cautelares de diversa intensidad junto o subsidiariamente a la concesión de medidas de protección autónomas que implementa el propio órgano persecutor penal.

La PUIR está construida en base a diversas dimensiones de riesgo. Cada una de ellas (5 en total) se compone de una serie de preguntas, las que se realizan a la víctima de forma secuencial. Dichas dimensiones corresponden a las características del delito denunciado, características del imputado, antecedentes históricos de violencia, características de la víctima y percepción de riesgo. Así, dentro de la dimensión referida a las “características del delito” denunciado se le pregunta a la víctima por golpes o intentos de golpes en los hechos que denuncian, lesiones, amenazas, uso de arma blanca, violencia sexual. En relación a la dimensión referida a las “características del imputado”, se le pregunta a la víctima, por ejemplo, si el sujeto tiene acceso a armas de fuego, si consume alcohol o drogas, si la golpea en dicho estado, si mantiene algún trastorno psiquiátrico, todas preguntas que apuntan a establecer la peligrosidad del sujeto denunciado. Es interesante señalar que esta dimensión se complementa con la obtención de información desde el propio sistema, como la existencia de causas previas en el registro del imputado, la existencia de otras denuncias o condenas por hechos similares (de carácter de VIF), o bien condenas o procesos pendientes por otros delitos. Puede apreciarse que en esta dimensión se contemplan preguntas que si bien se formulan sucesivamente (generando cada una de ellas una determinada puntuación), muchas consisten en una reiteración de la misma información. Así puede observarse, por ejemplo, respecto del uso de armas blancas o de fuego, y la eventual existencia de denuncias o condenas por dichos hechos. De este modo, es posible estimar que el puntaje final se verá directamente influido por una reiteración de información que ingresa por dos vías al sistema: aquella que aporta directamente la víctima y aquella que mantiene el propio sistema procesal penal. Por su parte, respecto de la dimensión referida a los “antecedentes históricos de la violencia”, se busca recabar información sobre la relación de violencia entre víctima e imputado en el transcurso del tiempo. Así las preguntas se refieren a la reincidencia, frecuencia, gravedad, amenazas previas o a la existencia de otras víctimas. Dentro de la dimensión referida a las “características de la víctima”, se recaban antecedentes

²⁶ BRAVO y BARRA (2022), pp. 1-28.

²⁷ De algún modo, esta descripción permite calificar a la PUIR como un algoritmo público, es decir, como un conjunto de pasos o instrucciones precisas diseñadas para realizar una tarea o resolver un problema específico. Asimismo, para el Laboratorio de Innovación Pública de la Escuela Gobierno de la Universidad Adolfo Ibáñez, la PUIR forma parte del Repositorio de Algoritmos Públicos, el cual contempla información sobre los sistemas de soporte o toma de decisiones automatizadas que han sido desarrollados y/o implementados por instituciones públicas en Chile.

²⁸ Fiscal Nacional, Oficio N.º 1032/2021 de 03 de diciembre de 2021. Instrucción general que imparte criterios de actuación en delitos de violencia de género y violencia intrafamiliar.

vinculados a su vulnerabilidad, tales como una condición de embarazo, de discapacidad o de dependencia económica respecto del imputado. Finalmente, la pauta contempla una dimensión referida a la “percepción de riesgo” que mantiene la propia víctima. Aquí, se recaban antecedentes referidos a su creencia respecto de futuras agresiones, o derechamente del riesgo que percibe en orden a ser víctima de un delito contra la vida, ella o sus familiares. Las dos preguntas que se formulan dentro de este ítem llaman la atención, en tanto recogen aspectos subjetivos de la vivencia de la víctima, sus creencias o percepciones respecto de lo que la persona denunciada podría hacer en el futuro.

El Ministerio Público considera que la evaluación automatizada del riesgo en las causas VIF es esencial para el desempeño de sus funciones y ha valorado positivamente la PUIR, pues “ha mostrado mejoría en la precisión de las evaluaciones basadas exclusivamente en el juicio clínico o en la pericia de un experto en contextos psiquiátricos, penitenciarios y jurídicos”.²⁹ A pesar de la importancia que se le asigna a la herramienta por parte de los operadores públicos, hay un marcado déficit de transparencia activa al respecto. Sin acceso adecuado a los documentos que han servido de base al diseño e implementación de la PUIR, de sus procesos de validación o de sus evaluaciones internas, es difícil estimar la importancia o centralidad que tiene la herramienta en el abordaje de las causas VIF. Que una herramienta sea importante para un organismo público no debería implicar su opacidad. Ante una solicitud de acceso a la información pública relativa a la PUIR realizada en 2019, el entonces Ministro del Interior y Seguridad Pública reconoció la inexistencia de un “documento físico” en que conste esta herramienta, pues se trata de un

“un programa computacional desarrollado en base a algoritmos matemáticos (...) En consecuencia, al ser la [PUIR] un programa computacional y cuyos algoritmos son la base fundamental del mismo, se hace necesario señalar que la información solicitada resulta sensible para la seguridad de la Nación, debido a que dicha Pauta es utilizada por la Policía de Investigaciones, Carabineros de Chile y el Ministerio Público [...]”.³⁰

¿En qué sentido el conocimiento público del modo en que esta herramienta asigna los puntajes y genera los diferentes niveles de riesgo podría representar “sensible para la seguridad de la Nación”? ¿Cómo se podría evaluar la importancia de esta herramienta en la persecución de los delitos y en la adopción de medidas de protección a las víctimas? ¿Cómo podemos evaluar si acaso la herramienta tiene un rendimiento con niveles adecuados de precisión o con razonables tasas de error para los diferentes grupos de mujeres? Considerando la falta de validación del instrumento, al no tener acceso al proceso de evaluación realizado por la Universidad Católica Silva Henríquez, y la falta de acceso a la información sobre las razones que motivan su diseño o al modo en que se jerarquizan y ponderan los puntajes de las diversas preguntas, es difícil o imposible contestar estas preguntas de manera adecuada.

²⁹ BRAVO y BARRA (2022), p.3.

³⁰ Subsecretaría del Interior, Ordinario N.º 18.311 de 28 de junio de 2019 que informa sobre solicitud de información N.º 1 AB001W0010259, según lo dispuesto en el inciso segundo del artículo 10 de la Ley N.º 20.285, sobre acceso a la información pública.

Además, como señalamos antes, en casos de riesgo medio o alto, las instrucciones del Fiscal Nacional, la autoridad superior del servicio, demandan la adopción de medidas cautelares.³¹ De acuerdo con este protocolo de actuación, cuando las fiscalías se enfrenten a una denuncia que califique como de riesgo alto, se deberá obtener alguna de las siguientes medidas cautelares: la detención del imputado, la prohibición de acercarse a la víctima, la salida inmediata del agresor del hogar común y/o la prohibición de porte y tenencia de armas de fuego o su incautación en el caso de mantenerse estas por el imputado. De esta forma, el resultado de la pauta obliga a la adopción de medidas cautelares que requieren la judicialización de la causa.

Más allá del problema que afecta a las víctimas, que ignoran el modo en que la PUIR aborda sus denuncias, el hecho de que en casos de riesgo alto haya una obligación de solicitar medidas cautelares de alta intensidad agrega un problema adicional. En concreto, ello pudiera resultar problemático para el derecho a la defensa y el debido proceso penal, pues no existe claridad alguna respecto de la forma en que se realiza la ponderación y puntuación de los resultados de las preguntas formuladas. La necesidad de obtener medidas cautelares obliga a la judicialización de la causa, lo que a su vez obliga a fijar un plazo de investigación de menor duración en tanto más gravosa sea la medida cautelar. Cabe preguntarse si efectivamente ello es una solución adecuada para todos los casos de VIF que maneja el sistema, como también si una investigación breve se ajusta a la complejidad que demandan estos casos respecto de la recolección de los medios de prueba necesarios. En este escenario, la existencia de un modelo automatizado de protección redundante en un sistema de gestión de causas también automatizado y breve, lo que en algunos casos pugnaría, sin lugar a duda, con las sanciones que en definitiva se impongan a los agresores en estos casos.

Asimismo, es posible sostener que algunos ítems de la PUIR duplican la información que se pretende recabar, con el consecuente aumento del puntaje, tendiendo así a incrementar los casos de riesgos alto. Esto último pone en cuestión la objetividad del mecanismo, pero además genera dudas en orden a la futura tramitación de la causa y la consideración de los intereses de la víctima. Así, nos parece que cabe ser cuidadosos respecto del modelo de tramitación gerencial o automatizado, pues ello podría poner en riesgo el derecho a la tutela judicial de las víctimas y, por cierto, al derecho a la igualdad de trato entre aquellas que denuncien causas similares que arrojen diversos niveles de riesgo. Del mismo modo, e insistiendo en la necesidad de conocer la validación de esta pauta y del modo de funcionamiento del mecanismo de puntuación, estamos ante una situación que afecta a diversos derechos fundamentales, tanto de las víctimas como de los imputados.

³¹ Entre otras medidas, estas instrucciones señalan que si en audiencia de control de detención o de formalización se presentaren casos evaluados con riesgo vital/alto, los fiscales deberán solicitar la prisión preventiva e informar a la URAVIT de todo caso en que dicha medida no haya sido concedida, según lo dispuesto en la Instrucción general que imparte criterios de actuación en delitos de violencia de género y violencia intrafamiliar contenida en el Oficio N.º 1032/2021 de 03 de diciembre de 2021.

2. Uso de IA en investigación de delitos VIF o violencia de género en el derecho comparado: el caso del VIOGEN

La automatización de los procesos de evaluación de riesgo es una tendencia global. En este escenario, la adopción de herramientas de IA aparece como una fase lógica del proceso. De acuerdo con un reportaje recientemente publicado por el New York Times, el caso español “ejemplifica cómo los gobiernos están recurriendo a algoritmos para tomar decisiones sociales, una tendencia global que se espera crezca con el auge de la inteligencia artificial.”³² Por ello, nos parece interesante incorporar una sección especial que estudie el caso del VIOGEN en España. Más allá de las críticas y polémicas que se han levantado en el último tiempo, VIOGEN es un sistema pionero en Europa que ha demostrado cierto grado de eficacia en la evaluación del riesgo y prevención de la violencia de género. Así, es un modelo probado que puede ser adaptado al contexto chileno. Por lo mismo, un análisis de VIOGEN permitirá identificar buenas prácticas, lo que facilitará la implementación de estrategias más robustas y efectivas en Chile. Finalmente, el enfoque de derecho comparado enriquece el marco normativo y técnico chileno al proporcionar perspectivas internacionales que pueden fortalecer la legislación y las políticas públicas locales en la lucha contra la VIF.

En el marco de la Ley Orgánica 1/2004 de Medidas de Protección Integral contra la Violencia de Género, vigente desde 2007,³³ España implementó el Sistema de Seguimiento Integral en los casos de Violencia de Género (VIOGEN) para reducir la violencia de género y proteger a las víctimas. VIOGEN evalúa el riesgo de que una mujer que ha denunciado vuelva a sufrir violencia a corto o medio plazo, permitiendo así una protección proporcional y oportuna. Dependiente de la Secretaría de Estado de Seguridad del Ministerio del Interior, VIOGEN utiliza una base de datos alimentada por valoraciones policiales a través de dos formularios: la Valoración Policial de Riesgo (VPR) y la Valoración Policial de Evolución de Riesgo (VPER).³⁴ Así, el sistema clasifica los niveles de riesgo como “No apreciado”, “bajo”, “medio”, “alto” y “extremo”, y, según ello, se implementan las medidas de protección establecidas en los anexos de la Instrucción SES 7/2016.³⁵ VIOGEN también permite la emisión de alertas tempranas y un seguimiento continuo de las víctimas.

El sistema VIOGEN es una aplicación informática/web integrada en la Red SARA (sistema de aplicaciones y redes para las administraciones), que almacena datos de denuncias policiales, resoluciones judiciales, información penitenciaria y antecedentes penales relacionados con la violencia de género. También incluye diversos datos personales y de servicios sociales y asistenciales para configurar los expedientes de las víctimas. En términos concretos, VIOGEN funciona como una plataforma que unifica o integra bases de datos fragmentadas de diversos servicios públicos y autoridades policiales, con el fin de filtrar y corregir la información para aplicar modelos automatizados en el tratamiento de estos datos.³⁶ De manera progresiva,

³² SATARIANO y TOLL PIFARRÉ (2024), *passim*.

³³ Red Europea de Prevención del Delito (2021), *passim*.

³⁴ GONZÁLEZ *et al.* (2018), pp. 55-64.

³⁵ ROA *et al.* (2023), p.360-390.

³⁶ GONZÁLEZ (2018), *passim*.

VIOGEN ha ido incorporando nuevas bases de datos y mejorando su capacidad de detectar patrones de violencia de género a partir de la vulnerabilidad o riesgo de los hijos menores, generando alertas automatizadas a los agentes policiales encargados de la protección de la víctima.³⁷ Una ventaja clave del sistema es la asignación automática de casos a la unidad policial correspondiente y la emisión de notificaciones automáticas sobre plazos y cambios de situación.³⁸ Esto estandariza la estimación policial de riesgo y las medidas de protección.

En el marco de la interacción entre el VPR y el VPER,³⁹ la frecuencia de monitorización varía según el nivel de riesgo inicial, de 72 horas para riesgo extremo a 60 días para riesgo bajo.⁴⁰ Los agentes pueden ajustar la valoración según su investigación, lo que incluye la opinión de profesionales de otras áreas, como médicos y psicólogos forenses para considerar factores adicionales. En 2017, había más de 42,300 usuarios habilitados, cada uno con un perfil específico y acceso limitado (con intervenciones auditables), de acuerdo con la Ley Orgánica de Protección de Datos.⁴¹ Además, todos los usuarios reciben capacitación continua a través de cursos obligatorios y voluntarios para mejorar la implementación del sistema.⁴²

A pesar de las ventajas que ha tenido el VIOGEN en la prevención de la violencia de género, la herramienta ha sido objeto de diversas críticas. Una de las principales críticas se dirige hacia el gran porcentaje de casos que el sistema automatizado declara como de riesgo “no apreciado”, que asciende a un 45% de las denuncias.⁴³ Este problema ha sido abordado en algunas sentencias judiciales, una de las cuales merece la pena analizar en detalle.⁴⁴ En 2020, la Audiencia Nacional de España conoció un caso de homicidio de dos mujeres, madre e hija, a manos del esposo de la primera. El caso dio cuenta de que la Guardia Civil correspondiente denegó la medida solicitada oportunamente por la víctima, considerando que la calificación elaborada por el modelo fue la de “riesgo bajo”.⁴⁵ El fallo ilustra algunas deficiencias estructurales del VIOGEN, como el hecho de que no fueron contemplados como parámetros de la investigación factores referidos a la violencia sufrida, ni comprobados los antecedentes del agresor, las circunstancias familiares, sociales (familia directa, hijos), económicas ni laborales (actividad laboral o profesional ni recursos propios)⁴⁶, ni los nuevos antecedentes aportados oportunamente por la víctima, todo lo cual impidió que se corrigiera la valoración inicial de riesgo. Como vemos, una de las críticas principales está dirigida a la forma de implementación y a la falta de capacitación de las policías y funcionarios encargados de interactuar con el sistema y realizar una apreciación del riesgo. Otro de los desafíos de implementación deriva del

³⁷ Ministerio de Igualdad España (2022).

³⁸ Ministerio de Igualdad España (2022).

³⁹ Instrucción 10/2007, España.

⁴⁰ GONZÁLEZ (2018), pp.83-102.

⁴¹ GONZÁLEZ (2018), pp.83-102.

⁴² De acuerdo con el “IX informe CEDAW”, publicado por la Convención sobre la Eliminación de Todas las Formas de Discriminación contra la Mujer en marzo de 2022.

⁴³ LÓPEZ DE ZUBIRÍA DÍAZ (2024), p. 83.

⁴⁴ Sentencia (Sala de los Contencioso-Administrativo, sección quinta) de 30 de septiembre de 2020. Sobre recurso 2187/2019.

⁴⁵ Sentencia (Sala de los Contencioso-Administrativo, sección quinta) de 30 de septiembre de 2020. Sobre recurso 2187/2019.

⁴⁶ Sentencia (Sala de los Contencioso-Administrativo, sección quinta) de 30 de septiembre de 2020. Sobre recurso 2187/2019.

hecho de que la recolección de datos está basada en respuestas de víctimas que están en un estado emocional alterado y que pueden no estar aptas para responder preguntas cerradas.⁴⁷ Diversos autores destacan la necesidad de apoyo profesional, legal y psicológico desde el inicio de la denuncia para captar mejor el riesgo real de la víctima.⁴⁸

Otras dificultades analizadas tanto por la doctrina como por la sociedad civil se relacionan con problemas que afectan, en general, a toda herramienta o algoritmo de IA. Uno de los grandes problemas de todo sistema automatizado de decisión deriva del denominado “sesgo de automatización”, que consiste básicamente en la predisposición de los seres humanos a no cuestionar los resultados de estos sistemas, aun cuando puede existir otro tipo de fuentes que sugieran un resultado distinto. En el caso del VIOGEN, diversos estudios hablan de que un 95% de los oficiales o profesionales a cargo se conforman con el nivel de riesgo ofrecido por el sistema.⁴⁹ Esto, a su vez, se traspa a las decisiones judiciales que deben aprobar las solicitudes de medidas de protección o cautelares que pueden decretarse, que también suelen tener cierto grado de deferencia con las apreciaciones de las policías u órganos persecutores. Otro sesgo importante que puede afectar a la implementación de este sistema es aquel relacionado con las bases de datos que sirven de entrenamiento y retroalimentación al algoritmo que están en la base del VIOGEN. En efecto, el algoritmo se comportará de mejor manera con aquellas víctimas que estén mejor representadas en esas bases de datos, generando un sesgo hacia el promedio. Como han destacado algunos autores, existe el riesgo de que “ciertas víctimas que no se inserten en los perfiles de estas herramientas y que, por tanto, en lugar de contar con herramientas que apoyen la prevención, nos encontremos con sistemas que desvirtúen el objetivo, excluyendo institucionalmente a mujeres víctimas por no acomodarse en el perfil de víctima que ha definido el sistema”.⁵⁰

Por otra parte, un problema que afecta a varios sistemas automatizados de evaluación del riesgo de víctimas de violencia es la falta de validación de los instrumentos o herramientas. De existir evaluaciones o auditorías que supervisen el monitoreo o desempeño de estos modelos, éstas suelen permanecer opacas para el público. En el caso del VIOGEN, en el año 2018 la fundación Éticas solicitó una auditoría interna al Ministerio Español, sin respuesta positiva, cuestión que los motivó a realizar una auditoría externa.⁵¹ El informe critica al sistema VIOGEN por su limitada accesibilidad a medidas de protección para mujeres con hijos, inmigrantes sin papeles y mujeres dependientes del agresor. Advierte al gobierno español sobre la necesidad de revisar el sistema, dado que 14 de 15 mujeres asesinadas habían sido clasificadas como de bajo riesgo.⁵² El informe también señala la falta de transparencia del algoritmo, la insuficiente formación policial, y la falta de participación de las propias víctimas tanto en el diseño como en la

⁴⁷ GONZÁLEZ *et al.* (2018), p. 116.

⁴⁸ Éticas (2022), p. 25.

⁴⁹ ZURITA (2014), p.827.

⁵⁰ LÓPEZ DE ZUBIRÍA DÍAZ (2024), p. 86.

⁵¹ Véase Éticas (2022). El informe destaca que solo 1 de cada 7 mujeres que busca ayuda recibe protección, ya que el sistema asigna a la mayoría el nivel de riesgo “no apreciado”, y solo un 3% de ellas son consideradas en nivel “medio”. Solo a partir de este nivel se otorga efectiva protección policial. Esta deficiencia es evidente, dado que entre 2003 y 2021 se registraron 71 homicidios de mujeres que previamente habían denunciado su situación.

⁵² Éticas (2022), p.21.

implementación del sistema. Además, revela que el 80% de las mujeres entrevistadas tuvo problemas con el sistema y solo el 35% conocía su nivel de riesgo, lo que demuestra un deficiente manejo informativo y posibles inexactitudes en la recopilación de datos.⁵³

3. El ASR

El ASR es un algoritmo público de evaluación del riesgo de las mujeres víctimas de violencia en relaciones de pareja. En tal calidad, se trata de un “sistema automatizado que utiliza algoritmos, simples o complejos, cuyo fin es el apoyo de la toma de decisiones o la automatización de decisiones en el sector público.”⁵⁴ En su Cuenta Pública de 2023, el Fiscal Nacional de Chile destacó “la futura implementación, en todas las Fiscalías Regionales, del Proyecto Algoritmo de Sugerencia de Riesgo, que nos permitirá contar con una situación de riesgo actualizada que alertará a los equipos si es necesario modificar las medidas de protección en favor de la víctima de violencia intrafamiliar.”⁵⁵ Actualmente, el ASR se encuentra en etapa de pilotaje en cinco Fiscalías Regionales (Arica, Valparaíso, O’Higgins, Los Lagos y Magallanes)⁵⁶. De acuerdo a la información recabada, el ASR ha sido utilizado en aproximadamente más de 8.000 casos conocidos por estas fiscalías, específicamente los delitos VIF contra mujeres.

3.1. Metodología

Considerando la falta de información pública que nos permita evaluar la importancia que este algoritmo público tendría para el órgano encargado de la persecución penal, decidimos adoptar un enfoque cualitativo que nos permitiera conocer, de primera fuente, el diseño e implementación o pilotaje de este sistema en diversas regiones del país. La metodología cualitativa se caracteriza por otorgar mayor flexibilidad y dinamismo al proceso de investigación producto de “la propia naturaleza del objeto de estudio, las realidades humanas”.⁵⁷ Esta metodología puede ser comprendida, a su vez, desde un enfoque socio-técnico, prevalente en el estudio de las nuevas tecnologías digitales.⁵⁸ Adoptar este enfoque supone considerar tanto los aspectos sociales como técnicos del diseño e impacto de las nuevas tecnologías digitales, que no existen en un vacío, sino que están profundamente integradas en contextos sociales, económicos y culturales.

Para la recolección de información, se realizaron entrevistas semiestructuradas y una solicitud de acceso a la información pública. Originalmente se planeó entrevistar a 9 funcionarios de 3 fiscalías regionales, pero finalmente se entrevistó a 3 funcionarios de una sola Fiscalía Regional entre noviembre de 2023 y enero de 2024, debido a la negativa del Ministerio Público para

⁵³ Éticas (2022), p.24.

⁵⁴ GobLab Universidad Adolfo Ibáñez (2023), p. 26.

⁵⁵ VALENCIA (2023), p.10.

⁵⁶ Durante el desarrollo de este trabajo, el piloto comenzó a utilizarse en todas las fiscalías regionales del Ministerio Público de Chile.

⁵⁷ GONZÁLEZ GIL (2009), p.122.

⁵⁸ HOLTON y ROSS (2021), *passim*.

incluir a las otras dos fiscalías.⁵⁹ En la Fiscalía seleccionada, 9 profesionales utilizan el ASR: 8 de la Unidad Regional de Atención a las Víctimas y Testigos (URAVIT) más la jefatura. También entrevistamos a un funcionario involucrado en las etapas previas al diseño y pilotaje del ASR; esta reunión no pudo ser grabada, por lo que solo se cuenta con las notas. Adicionalmente, a partir de una solicitud de acceso a la información pública realizada ante el Ministerio Público, concretamos una reunión telemática con la División de Atención de Víctimas y Testigos.⁶⁰ Si bien no se nos permitió grabar esa reunión, el representante de aquella unidad nos envió una presentación con diapositivas.⁶¹

En la respuesta a nuestra solicitud de acceso a la información, la Fiscalía Nacional nos informó que el proyecto de ASR dejó de estar operativo en diciembre del año 2022, y que sería reanudado en agosto de 2024, ya que en enero de este año iniciaría un proceso de licitación pública “con el fin de contar con asesoría especializada para poder implementar ajustes relevantes requeridos al algoritmo, al proceso de implementación y al aplicativo que se utilizará”⁶². Sin embargo, a partir de las entrevistas, constatamos que el ASR se siguió utilizando con posterioridad al mes de diciembre de 2023.

En febrero de 2024, la Fiscalía Nacional lanzó una licitación pública para la "Provisión de data center gestionado y servicios asociados para el Ministerio Público". Durante este breve periodo, las personas jurídicas interesadas podían inscribirse para obtener las bases de la licitación.⁶³ El procedimiento está regulado por el “Reglamento sobre procedimiento para licitación pública” del Ministerio Público, que incluye normas especiales para procesos tecnológicos complejos y críticos.⁶⁴ Este reglamento establece el principio de neutralidad tecnológica, prohibiendo preferencias por marcas específicas, salvo excepciones justificadas. Además, contempla un proceso previo con consulta a la industria (RFI) y criterios específicos para la elaboración de las bases de licitación y la evaluación de ofertas. En cuanto al manejo de información, considera la relevancia de los datos a los que accede el proveedor, los requerimientos normativos y la interoperabilidad e intraoperabilidad de los sistemas involucrados.⁶⁵ Más allá de una referencia específica a la necesidad de garantizar el cumplimiento de la Ley 19.628, no existe ninguna provisión concreta acerca del respeto y protección de los datos personales que puedan estar implicados en el uso de tecnologías desarrolladas por actores privados.

3.2 Las motivaciones y objetivos del ASR

⁵⁹ Al contactar a funcionarios de la segunda fiscalía seleccionada, se nos señaló que no contaban con el permiso de la División de atención a víctimas y testigos de la Fiscalía Nacional.

⁶⁰ Solicitud iniciada con fecha 5 de Julio de 2023, incluida en el folio SIAU N°19766 (véase anexo n°2 para el detalle de la solicitud).

⁶¹ La reunión virtual se realizó el día Jueves 11 de agosto de 2023.

⁶² Fiscalía Nacional, Carta DEN LT N° 936/2023, de 29 de diciembre de 2023.

⁶³ Si bien estas bases no están públicamente disponibles, pudimos constatar que se incluían servicios asociados a la evaluación e implementación del ASR.

⁶⁴ Texto aprobado por Resolución FN/MP N°1186/2022, de 9 de septiembre de 2022.

⁶⁵ Artículo 37 del Reglamento sobre procedimiento para licitación pública del Ministerio Público.

Considerando que cualquier algoritmo público pretende abordar un problema de interés colectivo, el ASR parte de la base que el problema de la violencia de género genera “víctimas fatales permanentemente y [que] se ha acrecentado también el número de femicidios frustrados.”⁶⁶ Además, el desarrollo del ASR considera tres factores principales: las limitaciones de la Pauta PUIR, principalmente porque es una herramienta estática y basada en el auto reporte; la falta de recursos humanos adecuados para monitorear, de manera constante, la situación de las víctimas; y, por último, el hecho de que si bien el Ministerio Público cuenta con gran cantidad de datos que son útiles para el análisis del riesgo de las víctimas, no tiene capacidades de analizarlos o procesarlos oportunamente. En palabras de la propia Fiscalía Nacional, el propósito del ASR es “[m]ejorar la información (calidad y oportunidad) de la probable situación de riesgo de una víctima, a fin de mejorar la toma de decisiones de quienes tienen la responsabilidad de diseñar e implementar las estrategias de protección.”⁶⁷

En el caso del ASR, las URAVIT, compuestas por profesionales como psicólogos, trabajadores sociales y abogados, deben diseñar estrategias de protección a víctimas y testigos de delitos, evaluando el riesgo al que están expuestos y aplicando medidas necesarias para su seguridad.⁶⁸ El ASR permite estimar la probabilidad de que una mujer víctima de violencia de pareja sufra un nuevo hecho de violencia grave, proporcionando información rápida y clara a fiscales y otros profesionales. Este sistema automatizado analiza datos en los aplicativos de la Fiscalía, ⁶⁹ calcula la probabilidad de un nuevo evento grave, y envía alertas al equipo URAVIT. Utiliza un algoritmo de IA que se actualiza con nueva información y se entrena mensualmente, con datos históricos desde 2015, aunque no se conoce quién se encarga de este proceso de entrenamiento.

En este escenario, el desarrollo de un algoritmo que incorpora técnicas de IA permite analizar casi instantáneamente grandes cantidades de información e ir actualizando el estado o situación de las víctimas de acuerdo con el cambio de la información ingresada o chequeada periódicamente. Si bien el ASR podría ser descrito, en concreto, como la mera integración de diversas bases de datos, en la práctica funciona como un sistema automatizado de decisión que evalúa el riesgo de las víctimas. Sus datos originales de entrenamiento están constituidos a partir de datos históricos de los casos de todas las víctimas de violencia de pareja desde el año 2015. A diferencia de la PUIR, y en consonancia con el VIOGEN, el ASR se (re)entrena mensualmente, para mejorar su capacidad de predicción, aunque no hemos tenido acceso a cuál unidad o personas están encargadas de aquello.

3.3 ¿Cómo funciona el ASR?

El funcionamiento del ASR considera 26 variables base (116 en total), extraídas de diferentes bases de datos del Ministerio Público, que son pertinentes para estimar o analizar el potencial

⁶⁶ Información obtenida a través de diapositivas proporcionadas por la DAVT.

⁶⁷ Información obtenida a través de diapositivas proporcionadas por la DAVT.

⁶⁸ Artículo 6, Código Procesal Penal.

⁶⁹ El Ministerio Público utiliza los servicios de *Azure*, la plataforma de *cloud computing* desarrollada por Microsoft, e incluye más de 200 productos y servicios que se almacenan y ejecutan en la nube.

riesgo de las víctimas ante la ocurrencia de un nuevo hecho de violencia grave. En relación con la víctima, se consideran: la existencia de causas VIF anteriores con el mismo imputado; la existencia de causas VIF anteriores con otros imputados; las respuestas y el resultado de la aplicación de la pauta PUIR; y el número de causas de delitos violentos sufridos por la misma víctima. En relación con el imputado, se consideran: la existencia de causas VIF tanto con la víctima como con otras parejas; el número de causas de delitos violentos; y, por último, el número de desacatos, esto es, de incumplimiento de medidas ordenadas por un tribunal.

El componente que más específicamente se puede calificar como IA es la existencia de un algoritmo de procesamiento de lenguaje natural que se aplica sobre el texto contenido en la denuncia, en la declaración de la víctima ante la Fiscalía o en los contactos o las solicitudes que realice ante la URAVIT u otra unidad. A pesar de que no tuvimos acceso a la tecnología específica, lo más probable es que se utilicen los servicios de lenguaje de *Azure (Microsoft)*,⁷⁰ que incluyen herramientas con características preconfiguradas, en este caso, por los profesionales de URAVIT. En este sentido, es plausible sostener que, por lo menos, se utilizan las siguientes configuraciones:

- Reconocimiento de entidades con nombre: es una característica preconfigurada que clasifica las entidades (palabras o frases) en textos no estructurado en varios grupos de categorías predefinidas o que pueden ir agregándose manualmente. Esto permite el procesamiento de los diversos cuerpos de textos de la víctima, por ejemplo, detectando todos los ejemplos de armas en un texto (por ej., “cuchillo” o “pistola”) o incorporando criterios o factores de riesgo asociados al miedo;
- Análisis de sentimiento y minería de opiniones: consiste en una colección de algoritmos de IA que permiten analizar preferencias o evaluaciones de la gente, opiniones positivas y negativas que se asocian a aspectos específicos del texto.

En cuanto a la interfaz (Fig. 1), esto es, el vínculo (diseño) que permite la interacción entre el ASR y los funcionarios que tienen acceso al mismo, esta se compone de dos plataformas: un *dashboard* que contiene información general (sólo visual), y un aplicativo funcional que permite revisar y modificar la información con más detalle. Este aplicativo tiene tres secciones distintas: una que contiene información general de la víctima y del imputado, es decir, las 26 variables base descritas anteriormente; otra que contiene un historial de alertas, con distintas asignaciones de riesgo para cada alerta, la que puede originarse por diversas vías, por ejemplo, por una nueva denuncia, en una solicitud de información y atención a usuarios (SIAU), o en una nueva causa grave que se abre contra el imputado; por último, una tercera sección contiene el análisis de los factores de riesgo que fueron detectados a través del procesamiento de lenguaje natural contenido en textos (denuncias, declaraciones, SIAU, etc.).

⁷⁰ Véase la página web <https://azure.microsoft.com/es-es/products/ai-services/ai-language>

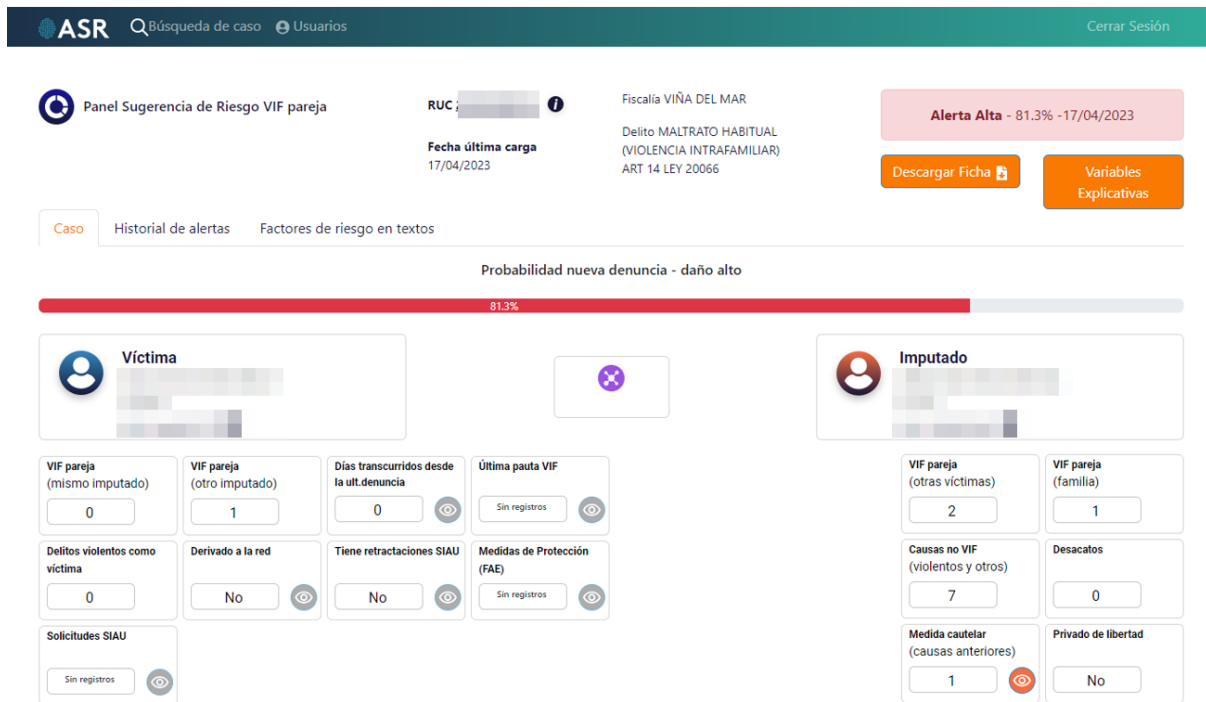


Figura 1

La segunda sección (Fig. 2) parece constituir un avance significativo con respecto a la información que puede extraerse de la pauta PUIR, y constituye un factor dinámico de la evaluación de riesgo (similar al VPER en el VIOGEN), que permite a los funcionarios de la URAVIT calificar el riesgo de acuerdo con el historial de alertas. Además, se despliegan hipervínculos con el historial de alertas de otras causas RUC del mismo imputado. Si van escalando en el tiempo, ello parece un indicio relevante para decretar o intensificar medidas de protección. De todos modos, esta sección supone la capacidad de los profesionales de URAVIT de analizar el detalle para clarificar dudas a través de la revisión de causas o análisis de información en profundidad.

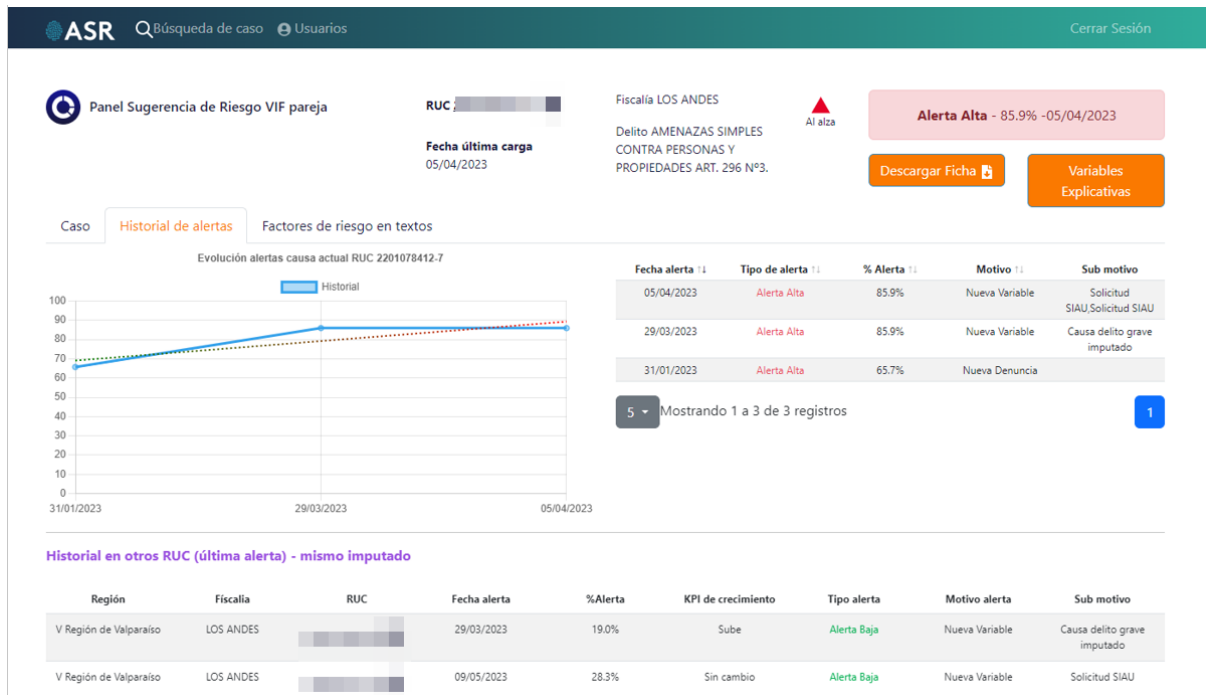


Figura 2

Todos los funcionarios entrevistados señalaron que el ASR es un instrumento que complementa la determinación del riesgo que ellos hacían en base a la información provista por la PUIR. Como ya señalamos, la gran ventaja del ASR con respecto a la PUIR es la capacidad de actualizar los niveles de riesgo, y evitar una representación de la realidad que queda detenida al momento de la denuncia. Una de las funcionarias destaca el mejoramiento de las capacidades de la Fiscalía con respecto a la PUIR:

“el algoritmo se va complementando con información de la historia, digamos, del proceso, entonces claro a lo mejor si una pauta arroja riesgo alto al minuto uno de la denuncia es probable que al mes siguiente ese nivel de riesgo haya cambiado por esas circunstancias, incluso por la implementación de medidas de protección, entonces lo vemos como un complemento que es valioso” (Funcionario 2).

Todos los entrevistados señalaron que el algoritmo es una herramienta cualitativamente mejor que la PUIR. Las razones de esta percepción se dan producto de la integración automatizada de bases de datos que antes se hacía manualmente:

“La idea es que toda la información que se aporta a través del parte policial o la posible, no necesariamente, pero la posible y deseable aplicación de la pauta PUIR, más toda la información que se aporte al sistema informático de la fiscalía se vaya integrando, incorporando y da como resultado una predicción de posibilidad de sufrir una nueva agresión y eso se va informando al equipo básicamente de URAVIT, de la unidad de atención de víctimas y testigos para efectos de evaluar y gestionar medidas de protección para esa víctima” (funcionaria 2).

Además, se trata de una herramienta que entrega valiosa información a los funcionarios de URAVIT. Como señala otro de los funcionarios:

“Es una herramienta super buena (...) este algoritmo puede hacer el análisis de información rápida, oportuna (...) indicándonos precisamente el riesgo, el nivel de riesgo, también de daño eventualmente, en una víctima de VIF, digamos, hay distintos daños, tanto los psicológicos, social, económico, entonces reúne elementos que nos arrojan un indicador de daño y eso, a mí por lo menos, me ha ahorrado un montón de tiempo (...)” (Funcionario 3).

Para los funcionarios, el uso del ASR permite el establecimiento de una estrategia de protección autónoma y, desde el punto de vista jurídico, la posibilidad del equipo de URAVIT de emitir informes a los fiscales para que puedan solicitar medidas dependiendo de la gravedad del caso:

“Es una herramienta que se suma a la pauta de riesgo y que aporta también en estos informes que el fiscal presenta en una causa por ejemplo para que se decreten medidas cautelares, entonces como te digo es una herramienta que yo siento que es bastante positiva (...) es una información súper relevante para nosotros” (Funcionario 3).

3.4 El desempeño del ASR en la práctica

En cuanto a las métricas de desempeño del ASR, hemos tenido acceso a una breve tabla que da cuenta del testeo que se hizo con casos reales, justamente para dar cuenta de la precisión y otras métricas con que se evalúa el rendimiento de los algoritmos. Para ello, es útil considerar la siguiente tabla (Fig. 3), incluida en la presentación de la Fiscalía Nacional, y que contiene un “panel de métricas de desempeño”:

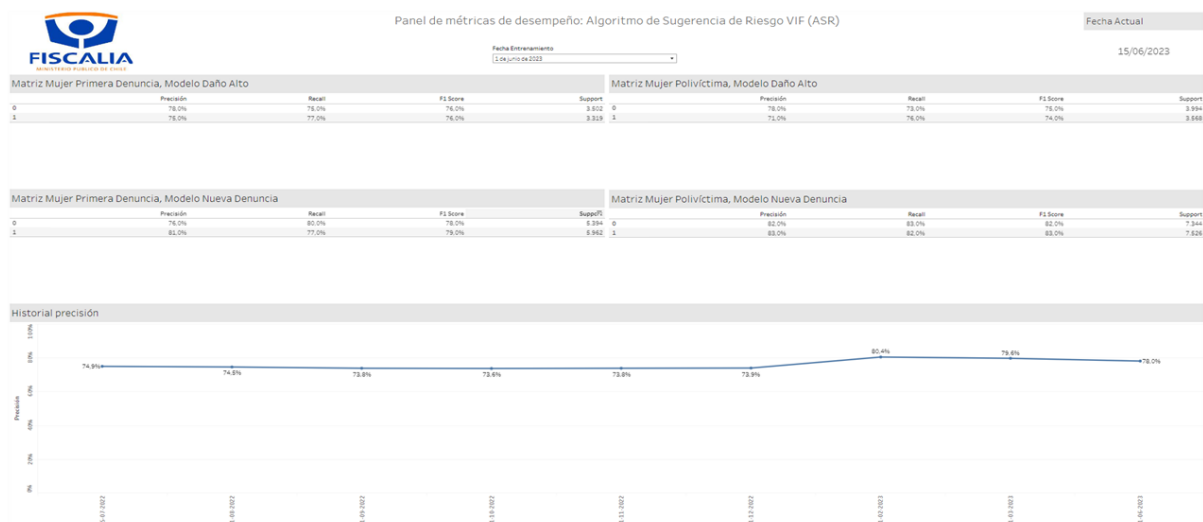


Figura 3

La imagen ilustra el funcionamiento de tablas de clasificadores de datos en cuatro modelos diferentes (Mujer Primera Denuncia, Modelo Daño Alto; Mujer Primera Denuncia, Modelo

Nueva Denuncia; Mujer Polivíctima, Modelo Daño Alto; Mujer Polivíctima, Modelo Nueva Denuncia). Para cada modelo, se muestran tres métricas para dos clases (0 y 1, que son básicamente las predicciones y los casos reales), así como el soporte (*Support*), que indica el número de casos en cada clase. Las métricas son las siguientes:

- “Precisión”: define la proporción de verdaderos positivos (TP) sobre el total de predicciones positivas, es decir, mide cuántas de las instancias que el modelo predijo como positivas realmente lo son;
- “Recall”: define la proporción de verdaderos positivos (TP) sobre el total de instancias que son realmente positivas (TP + FN), es decir, mide cuántas de las instancias realmente positivas fueron correctamente identificadas por el modelo; y
- “F1 Score”: es una métrica de rendimiento utilizada para evaluar modelos de clasificación, especialmente cuando se tienen clases desbalanceadas, es decir, combina la precisión y el *recall* en una sola medida utilizando la media armónica de ambas.

Además, el gráfico “historial de precisión” muestra la evolución de la precisión del modelo a lo largo del tiempo.

A partir de estos gráficos, podemos concluir lo siguiente: 1) los modelos tienen un desempeño razonablemente bueno, con métricas de precisión, *recall* y F1 score en el rango del 70% al 80%; 2) los modelos de "Mujer Polivíctima, Nueva Denuncia" (es decir, de mujeres que ya tienen denuncias anteriores en otras causas) parecen tener las mejores métricas en general, con valores de precisión y *recall* superiores al 80%; 3) el historial de precisión muestra que el rendimiento del modelo ha sido bastante estable, con una ligera mejora en los últimos meses; 4) si bien las métricas de precisión y *recall* parecen bastante razonables, el problema es que estamos lidiando con casos en que, si bien parece no ser problemático equivocarse por exceso, las tasas de error para los agresores pueden derivar en medidas de protección o cautelares que afectan sus libertades. De acuerdo con esto último, sería útil conocer el modo en que se están evaluando estas métricas. En lo que sigue, evaluaremos el ASR de acuerdo a dos derechos fundamentales que nos permitirán la adecuación de este algoritmo público con estándares de derechos humanos que es posible identificar y aplicar en este caso.

4. El ASR y los derechos fundamentales en la era de la intermediación algorítmica de las relaciones públicas

En esta sección, evaluaremos críticamente el diseño y el funcionamiento del ASR, de acuerdo con la información que hemos recolectado. Para ello, usamos como estándar el modo en que los derechos fundamentales se pueden adaptar a los desafíos que plantea la intermediación algorítmica entre el Estado y los ciudadanos al sistema procesal penal. Si bien estos estándares podrían enmarcarse a través del análisis del derecho al debido proceso, con todas las implicancias y garantías asociadas al mismo, en esta sección analizaremos el potencial impacto que el ASR tiene sobre dos derechos fundamentales: el derecho fundamental a la protección de datos personales y el derecho a la igualdad y no discriminación. De algún modo, el respeto y

protección de estos derechos fundamentales contribuye al respeto y protección del debido proceso, u honran la idea de que los procesos penales deben realizarse conforme a los derechos fundamentales. El ejercicio que haremos no consiste estrictamente en la realización de una auditoría algorítmica, aunque varias de las cuestiones que analizaremos podrían perfectamente enmarcarse en una. Varias de las cuestiones analizadas en secciones anteriores se corresponden con la etapa de mapeo conceptual, en la cual se presenta la identificación de los datos de identidad del algoritmo público, como la fecha de creación del algoritmo, los datos sobre la arquitectura básica, el marco metodológico y explicación de la metodología utilizada para definir el modelo o información sobre las bases de datos utilizadas tanto para el desarrollo como el entrenamiento del algoritmo.

Según el manual elaborado por *Ethicas Consulting*, las auditorías algorítmicas deben seguir cuatro principios clave.⁷¹ Primero, el cumplimiento ético y legal exige que el algoritmo respete el marco jurídico y los derechos fundamentales. Segundo, la deseabilidad implica que el sistema debe ser explicable, preciso, replicable, transparente, justo y eficiente, sin discriminación ni sesgos. Tercero, la aceptabilidad social requiere que el algoritmo sea comprensible, controlable, sostenible y beneficioso, con información clara para que la ciudadanía pueda evaluar sus efectos. Cuarto, la protección y gestión de datos asegura la calidad, actualización, fiabilidad y transparencia a lo largo del ciclo de vida del algoritmo. En este escenario, creemos que el análisis desde una perspectiva de derechos fundamentales puede enmarcarse como una operativización de estos principios rectores.

4.1. El derecho fundamental a la protección de datos personales

En el derecho comparado, y sobre todo con el desarrollo del derecho europeo, el derecho a la protección de datos personales ha adquirido el carácter de derecho fundamental. En nuestro ordenamiento jurídico, y desde 2018, el derecho a la protección de datos personales fue reconocido como derecho constitucional a través de la ley 21.096, que modificó el artículo 19 n° 4 de la Constitución. Lo que antes se había reconstruido doctrinariamente y en diversa jurisprudencia a partir de una lectura amplia del derecho a la privacidad, ahora se reconoce con un carácter independiente, esperando que su configuración legislativa le otorgue contornos definitivos a este nuevo derecho.⁷² En virtud de este derecho, es posible realizar diversos análisis del ASR, sobre todo considerando la pronta entrada en vigencia de una normativa actualizada sobre la materia, que además crea la Agencia de Protección de Datos Personales.⁷³

De acuerdo con la clásica consagración de los derechos ARCO, es plausible sostener que las víctimas cuyos casos han sido procesados por el pilotaje del ASR no han sido informadas de la recolección de sus datos personales o de otro tipo, de la finalidad específica que cumplen en el marco de este sistema automatizado de decisión, del grado de automatización del

⁷¹ Se utiliza como referencia la guía de auditoría algorítmica del equipo *Ethicas Research and Consulting SL*, <https://eticas.ai/adversarial-algorithmic-auditing-guide/>

⁷² CONTRERAS (2020), pp. 87-120.

⁷³ La ley 19.628 ha sido recientemente actualizada por la aprobación de un proyecto de ley que la reforma. Al momento de editar este artículo, el proyecto se encontraba en el trámite de control preventivo ante el Tribunal Constitucional. La nueva normativa entrará en vigencia el 01 de diciembre de 2026.

procedimiento, de los funcionarios que tienen acceso a ellos, o de las eventuales posibilidades de ejercer el derecho a rectificación u oposición en casos determinados. Como reconoce uno de los funcionarios, todo lo relacionado con el ASR es “un trabajo hasta el momento interno en el caso nuestro (...), porque además queremos que esto esté funcionando ya no de manera intermitente (...) que ya esté más consolidado para poder trabajarlo más directamente con los equipos jurídicos, más allá de nuestros informes, nuestros reportes etcétera, y además también con la víctima.” (Funcionario 1). Hasta el momento, y de acuerdo con las directrices internas emitidas por el Ministerio Público, la única obligación que recae sobre los fiscales es la de notificar y explicar a la víctima el modo en que funciona la PUIR: sólo se avisa “cuando se le aplica la pauta a la víctima, eso sí que se le explica una pauta que tiene digamos esta función, que es una fotografía del momento, eso se le explica al momento en que ella accede a contestar, por lo menos cuando la aplicamos nosotros como fiscalía” (Funcionario 2). De acuerdo con un oficio del Fiscal Nacional, que entrega instrucciones generales sobre los criterios de actuación que deben regir en materia de violencia de género, el deber de protección a las víctimas de violencia de género obliga a aplicar estándares de buen trato, “acceso a la información, participación y protección”.⁷⁴ A pesar de ello, todos los funcionarios entrevistados desconocen si acaso existió alguna precaución o medida en las etapas de diseño o implementación del ASR. Tampoco existe una normativa sobre el pilotaje de este algoritmo, que se ha aplicado hasta el momento para el procesamiento de más de 8.000 casos.

El problema, como vemos, no se trata sólo de la víctima, sino también del procesamiento de datos personales del imputado, quien potencialmente puede ver afectado su derecho a la libertad personal u otros. Otro funcionario reconoce que el desconocimiento del ASR también se predice respecto del imputado: “[e]specíficamente yo desconozco que estas personas (imputados) sepan, pero en todo caso el algoritmo lo que hace es lo que podríamos estar haciendo nosotros, y esa información que está en un sistema interno de la fiscalía es un sistema en base a las causas que existen previas, eso va quedando y eso yo creo que, los imputados, no sé si todos lo sabrán” (Funcionario 3).

Además, vale la pena señalar la centralidad que en estos casos adquiere el denominado derecho a oponerse al tratamiento automatizado de datos personales. La normativa de protección de datos recientemente aprobada incluye una articulación que intenta abordar los problemas que han surgido a propósito de la interpretación de tal derecho en el marco del Reglamento General de Protección de Datos (RGPD) de la Unión Europea.⁷⁵ El artículo 8 bis de la nueva ley 19.628, que se titula “Decisiones individuales automatizadas, incluida la elaboración de perfiles”, señala:

“El titular de datos tiene derecho a oponerse y a no ser objeto de decisiones basadas en el tratamiento automatizado de sus datos personales, incluida la elaboración de perfiles, que

⁷⁴ Oficio FN N° 1032/2021, p. 26.

⁷⁵ Artículo 22.1 del RGPD: “Todo interesado tendrá derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos en él o le afecte significativamente de modo similar.”

produzca efectos jurídicos en él o le afecte significativamente. (...) En todos los casos de decisiones basadas en el tratamiento automatizado de datos personales (...), el responsable deberá adoptar las medidas necesarias para asegurar los derechos, libertades del titular, su derecho a la información y transparencia, el derecho a obtener una explicación, la intervención humana, a expresar su punto de vista y a solicitar la revisión de la decisión.”

De acuerdo con Contreras y Viollier, la redacción de este artículo supone “un injerto jurídico modificado y mejorado” de los problemas que ha tenido el artículo 22 del RGPD en Europa.⁷⁶ En efecto, de la lectura del artículo se puede ilustrar el esfuerzo del legislador chileno por hacerse cargo de los problemas doctrinarios y prácticos que se han generado en el derecho europeo. Primero, se evitan los problemas que, en la práctica europea, han permitido hipótesis que permiten no cumplir con los requisitos que derivan de este derecho incorporando participación humana meramente nominal o simbólica. En concreto, al eliminar el término “únicamente”, el legislador pretende evitar situaciones que formalmente caerían dentro del marco literal de la ley, pero que defraudan su espíritu, que es otorgar protección a los derechos o intereses de las personas sujetas a procesos automatizados de decisión. En segundo lugar, el articulado propuesto en el caso chileno aclara que no sólo existe un derecho de oposición, que puede o no ser activado por el titular, sino un “derecho a no ser objeto de decisiones basadas en el tratamiento automatizado de datos personales”. Para algunos comentaristas, esta redacción permitiría catalogar esta norma como una prohibición antes que como un derecho, cuestión fundamental para un contexto en que hay grandes asimetrías de información y poder.⁷⁷ En otras palabras, las obligaciones para el responsable del tratamiento de datos deben cumplirse sin necesidad de que el titular active el derecho de oposición. En tercer lugar, la propuesta de redacción del legislador chileno incluye explícitamente el derecho a una explicación, cuestión que en Europa ha sido reconocida únicamente en el considerando 71 del GDPR, esto es, en aquellas partes que no serían parte del derecho vinculante. Si bien el caso chileno no aclara si se trata de un derecho a la explicación de carácter abstracto y ex ante, o de si una explicación concreta y ex post acerca de la decisión automatizada que afecta los intereses del titular, es una cuestión doctrinaria y práctica que perfectamente pueden ir desarrollando tanto la potestad de interpretación administrativa que tendrá la nueva Agencia de Protección de Datos o la jurisprudencia judicial que deba abordar este asunto.

Considerando que el marco normativo será mucho más exigente para los responsables del tratamiento de datos, es plausible sostener que, si el objetivo del Ministerio Público es lograr que el ASR esté plenamente operativo, deberá realizar diversos ajustes.⁷⁸ A pesar de que el ASR, tal como su nombre indica, es un algoritmo de “sugerencia” que no reemplaza el juicio experto de los profesionales de la URAVIT, es evidente que la carga de trabajo y el denominado “sesgo de automatización” podrían eventualmente generar un riesgo de que la intervención humana sea de carácter meramente nominal o simbólica, tal como se ha criticado en el caso del

⁷⁶ Pablo Contreras y Pablo Viollier, “Derecho a una explicación y regulación de las decisiones automatizadas: contrapunto entre el Reglamento General de Protección de Datos y la reforma chilena en protección de datos personales” (Borrador en poder de las autoras).

⁷⁷ VIOLLIER y FISCHER (2023), pp. 151-169

⁷⁸ Según el artículo 1° transitorio de la nueva ley 19.628, hay un período de vacancia de 2 años, contados desde la publicación en el Diario Oficial.

VIOGEN explicado antes. De acuerdo con los datos que tiene el Centro de Estudios y Análisis del Delito (CEAD), sólo el 2022 los delitos VIF contabilizaron un total de 6.276 denuncias en la Fiscalía Regional que estudiamos, constituyendo un 30.3% del total de denuncias por delitos violentos.⁷⁹ Considerando que aproximadamente una cuarta parte de las denuncias es catalogada con “riesgo alto”,⁸⁰ es posible sostener que 8 funcionarios de URAVIT, que tienen diversas funciones críticas para el desempeño del Ministerio Público, tuvieron sólo un pequeño porcentaje de tiempo para dedicarle a los casos de riesgo alto, tanto diseñando medidas de protección personalizadas y adecuadas a la víctima como solicitando restricciones a los derechos de los imputados. En este escenario, es posible sostener que si el ASR alcanza niveles suficientes de precisión, existe el riesgo de que los funcionarios incurran en el denominado “sesgo de automatización”. Este sesgo, que ha sido demostrado en diversos estudios psicológicos, se refiere a una propensión o tendencia de los seres humanos a considerar como correcto el resultado de los sistemas automatizados de decisión, incluso en caso de que existan señales de advertencia o de información contradictoria provista por otras fuentes.⁸¹ Según las entrevistas que realizamos, los funcionarios de la URAVIT toman ciertos resguardos al implementar el ASR: por ejemplo, en caso de cualquier inconsistencia entre la pauta PUIR y el ASR, o en caso que existan dudas que puedan sugerir la existencia de un riesgo alto, se opta por considerar el caso como de alto riesgo.

Para los funcionarios de la Fiscalía Regional seleccionada, el ASR, al igual que la PUIR, es un instrumento de riesgo que puede arrojar resultados que no necesariamente coinciden con la experticia del profesional de URAVIT. De acuerdo con lo que nos informaron los entrevistados, la División de Atención a Víctimas y Testigos ha solicitado a las respectivas URAVIT regionales que informen sobre las discrepancias generadas entre lo que determina el algoritmo y lo que resulta del juicio profesional. La idea de esto es ir reflejando, de la forma más precisa posible, el nivel de riesgo concreto de la víctima, e ir retroalimentando y mejorando la precisión del ASR.⁸² Además, los entrevistados confirmaron la cautela con la que analizan o interpretan los resultados del algoritmo:

“por ejemplo, si había una predicción de un daño alto y la verdad que nosotros estimamos como unidad que en realidad era bajo, es importante reportar este tipo de posibles inconsistencias, lo mismo que nos puede ocurrir al revés digamos: que puede que no nos llegue ninguna alerta de una causa que sí efectivamente nosotros pensemos o estimemos que puede tener un daño alto” (Funcionario 2).

Además, los funcionarios confirmaron la aplicación práctica de un criterio que no consta en los protocolos de actuación contenidos en el respectivo instructivo de Fiscalía. URAVIT está obligada a intervenir en todos los casos de una PUIR de riesgo alto o en caso de una alerta alta del ASR:

⁷⁹ Estadísticas extraídas del portal del CEAD, <https://cead.spd.gov.cl/estadisticas-delictuales/>

⁸⁰ Información obtenida a través de diapositivas proporcionadas por la DAVT.

⁸¹ CUMMINGS (2006), pp. 23-31.

⁸² Funcionarios 1, 2 y 3

“...en el caso de un algoritmo alto, aun cuando tengamos una pauta de riesgo bajo, vamos a intervenir sí o sí, aun cuando después despejemos que en realidad ese riesgo a lo mejor no era tan alto. O sea los dos nos van a indicar la intervención obligatoria nuestra y la profundidad de esa intervención va a estar determinada por la interacción que tengamos con la víctima, pero siempre va a ser una alerta para intervenir” (Funcionario 3).

Como vemos, y tal como lo pronostican diversos estudios acerca del sesgo de automatización en las burocracias públicas, cuando los funcionarios tienen conciencia del riesgo que implican los sistemas automatizados, y tienen espacios para contrastar otras fuentes posibles de riesgo, la sinergia o interacción que se produce entre humano y máquina parece ser productiva.⁸³ A diferencia del caso del VIOGEN, en que más de 40.000 funcionarios tienen acceso al sistema, el ASR parece estar muy consciente de los eventuales errores y de la necesidad de mantener un rol fundamental para el juicio experto.

Otra de las cuestiones relevantes se relaciona con la explicabilidad de los resultados del ASR. Como ya dijimos, la explicabilidad en sistemas automatizados de decisión puede referirse a la comprensión general de la lógica y funcionamiento del sistema, cuestión que puede garantizarse ex ante (es decir, antes de que opere el sistema), o a la comprensión específica de cómo el algoritmo tomó una decisión en un caso concreto, cuestión que necesariamente debe garantizarse ex post.⁸⁴ La primera proporciona una visión abstracta y, en ocasiones técnica, que puede ser útil tanto para expertos como para usuarios, mientras que la segunda ofrece detalles específicos para asegurar transparencia y justicia en situaciones individuales.

Con respecto a la primera comprensión de la explicabilidad, una de las primeras cuestiones que notamos es que no sólo las víctimas desconocen el funcionamiento o la lógica del ASR, sino también los funcionarios. Aun así, valoran la plataforma y la interfaz del ASR, pues les permite acceder a los documentos de manera rápida y sencilla. De algún modo, lo que más valoran los funcionarios no es tanto el componente de IA del ASR, sino su capacidad de filtrar información que está disponible en grandes bases de datos de la propia Fiscalía. Además, los funcionarios asumen que su juicio o experticia profesional es fundamental para interpretar los resultados, de modo que niegan cualquier intento por reemplazar el juicio profesional por un sistema enteramente automatizado. Un funcionario nos señaló lo siguiente:

“no entiendo muy bien cómo se hacen las ponderaciones, uno lo que hace es trabajar con la víctima y hablar con ella y uno entiende, digamos, analiza otro tipo de factores, yo no sé si algún día eso podrá llegar a una cierta perfección, yo creo que no, yo creo que el tema del ojo experto de un profesional nunca va a poder ser suplido totalmente por la inteligencia artificial, pero eso es una visión que a lo mejor es mía, ni siquiera me atrevo a comprometer” (funcionario 1).

⁸³ ALON-BARKAT y BUSUIOC (2023), pp. 153-169

⁸⁴ KAMINSKI (2021), pp. 278-299.

Otro funcionario también admite su desconocimiento sobre los tecnicismos involucrados en el diseño y funcionamiento del ASR, pero admite que a través de un proceso de refinamiento ello podría ir mejorando la utilización de IA en la gestión del riesgo de violencia contra la mujer: “en el fondo yo creo que a través de afinar esas ponderaciones por quienes sí saben de inteligencia artificial puede mejorar esas brechas, que en todo caso tampoco es que sean grandes brechas, existen, pero no son tan tremendas” (Funcionario 2).

En relación con la segunda comprensión de la explicabilidad, entendida ahora como la explicación del modo en que el algoritmo consideró las diversas variables para generar un resultado en un caso concreto, hay diversas dudas en relación al funcionamiento del ASR. Por una parte, tal como nos señaló una funcionaria que participó de las etapas iniciales del diseño del ASR, uno de los objetivos principales de este algoritmo público era de proveer razones que pudieran explicar una alerta alta o un aumento evolutivo del riesgo en el marco de una causa, y de este modo servir de apoyo ante cualquier solicitud que la Fiscalía deba hacer ante un juez de garantía para afectar derechos del imputado.

Como se puede ver en la Figura 4, la interfaz del ASR les permite a los funcionarios autorizados para hacer uso de este algoritmo acceder a la información sobre las “variables explicativas”. Si bien hicimos algunas consultas relativas al modo en que el ASR puede explicar de manera precisa el modo en que consideró las diferentes variables en un caso concreto, los funcionarios URUVIT entrevistados manifiestan no tener un conocimiento detallado de esto.

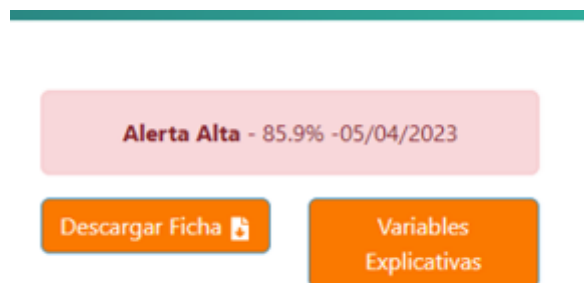


Figura 4

Los funcionarios de la URUVIT entrevistados también desconocen si acaso el pilotaje del ASR incluyó alguna evaluación prospectiva de los riesgos que implica, ya sea para la protección de datos personales u otros derechos fundamentales. Es importante recordar que el artículo 15ter de la nueva normativa de protección de datos incluye una obligación específica en este sentido:

“Cuando sea probable que un tipo de tratamiento, por su naturaleza, alcance, contexto, tecnología utilizada o fines, se pueda producir un alto riesgo para los derechos de las personas titulares de los datos personales, el responsable del tratamiento deberá realizar, previo al inicio de las operaciones del tratamiento, una evaluación del impacto en

protección de datos personales. La evaluación de impacto se requerirá siempre en casos de:
a) Evaluación sistemática y exhaustiva de aspectos personales de los titulares de datos, basadas en tratamiento o decisiones automatizadas, como la elaboración de perfiles, y que produzcan en ellos efectos jurídicos significativos. b) Tratamiento masivo de datos o gran escala. c) Tratamiento que implique observación o monitoreo sistemático de una zona de acceso público. d) Tratamiento de datos sensibles y especialmente protegidos, en las hipótesis de excepción del consentimiento.”

De la lectura de esta norma, es evidente que cualquier implementación del ASR, incluso en etapa de pilotaje, debería cumplir con este requisito, que debe ser materializado en un tiempo “previo al inicio de las operaciones del tratamiento”. Es importante resaltar que la evaluación de impacto debe incluir un análisis del riesgo no sólo sobre los datos personales de las personas, sino sobre todos sus derechos, tal como se deriva de una interpretación de los artículos 1 y 15ter de la nueva normativa.⁸⁵ Además de establecer hipótesis de evaluaciones de impacto forzosas que no se relacionan necesariamente con el impacto sobre los derechos fundamentales de las personas, la norma es relativamente amplia al señalar que esta obligación se puede gatillar dependiendo de la “naturaleza, alcance, contexto o tecnología utilizada o fines” del caso concreto. Si bien existen dudas y cierto grado de incertidumbre con respecto al alcance de esta obligación para los sujetos privados,⁸⁶ no existen dudas acerca de la necesidad de que los órganos públicos que sean responsables del tratamiento de datos cumplan con realizar una evaluación de impacto.⁸⁷ Toda evaluación de impacto debe incluir la notificación previa a las comunidades afectadas para discutir la pertinencia de un sistema automatizado de decisión y asegurar la participación de estas personas. Debe abordar la descripción del tratamiento, su finalidad, la necesidad y proporcionalidad del procesamiento, la evaluación de riesgos y medidas de mitigación. La nueva Agencia de Protección de Datos Personales jugará un rol clave al establecer una lista de operaciones que requieran evaluación de impacto y ofrecer orientaciones mínimas.⁸⁸ La Fiscalía, como responsable del tratamiento, puede consultar a la Agencia si el procesamiento resulta de alto riesgo, buscando recomendaciones adecuadas.

4.2. El derecho fundamental a la igualdad y no discriminación algorítmica

Tanto la normativa sobre protección de datos como la regulación propuesta para el uso de la IA consideran el respeto y protección del derecho a la igualdad y no discriminación como una cuestión fundamental.⁸⁹ Además, como se destaca en la Política Nacional de IA (PNIA actualizada), “[l]os sistemas de IA se desarrollarán y se utilizarán incluyendo a diversos agentes y promoviendo la igualdad de acceso, la igualdad de género y la diversidad cultural, evitando al mismo tiempo los efectos discriminatorios y los sesgos prohibidos por el ordenamiento jurídico.”⁹⁰ Esta Política, que sirve de orientación al despliegue de toda la actividad estatal sobre

⁸⁵ Artículo 1 inciso 2° de la modificación a la ley 19.628.

⁸⁶ JANSEEN (2020), pp. 76-106.

⁸⁷ REISMAN *et al.* (2018).

⁸⁸ Artículo 15ter inciso 3° de la nueva ley 19.628.

⁸⁹ Boletín 16821-19, que regula los sistemas de inteligencia artificial.

⁹⁰ PNIA actualizada al 2024

la materia, nos permite distinguir algunas de las causas o formas en que la discriminación puede producirse en la era digital. Una comprensión del modo en que la discriminación puede manifestarse o configurarse en diversas formas o articulaciones nos permitirá diseñar herramientas institucionales adecuadas para su prevención y combate.

Las causas de la discriminación en la era digital pueden estar relacionadas con los prejuicios culturales de quienes diseñan los sistemas tecnológicos. Sin embargo, estos prejuicios no son la única razón por la cual los sistemas pueden tener sesgos o impactos discriminatorios. Aunque no exista animadversión o estereotipos en los diseñadores o programadores, los sistemas “inteligentes” pueden todavía causar discriminación. Durante el diseño, testeo, evaluación, implementación o ejecución de estas tecnologías, siempre hay decisiones humanas implicadas. Lo más preocupante es que incluso sin intenciones discriminatorias, estos sistemas pueden afectar negativamente a personas pertenecientes a categorías protegidas por leyes antidiscriminatorias. Siguiendo una clasificación planteada por Ginés y Fabrellas, según la cual existen cuatro tipos o formas de discriminación algorítmica,⁹¹ en lo que sigue explicaremos estas distintas manifestaciones realizando comentarios acerca del modo en que el ASR podría abordarlas. La discriminación algorítmica puede manifestarse de diversas maneras, afectando a individuos y grupos de forma directa o indirecta. A continuación, se presentan las tres principales formas de discriminación algorítmica y sus causas:⁹²

4.2.1. Diseño de Variables del Algoritmo

Los algoritmos pueden discriminar por el diseño de las variables o criterios que determinan el objetivo principal del mismo.⁹³ Así, por ejemplo, un algoritmo de reclutamiento y selección de personal tiene como objetivo principal buscar “buenos candidatos” que posteriormente sean “buenos trabajadores”. Si los criterios que determinan qué cuenta como un buen candidato o trabajador incluyen cuestiones como la experiencia ininterrumpida en trabajos previos, es probable que ese criterio impacte desproporcionadamente en el número de hombres y mujeres que formarán parte de ese grupo. En el caso del ASR, el objetivo principal es medir el riesgo de que las víctimas de VIF vuelvan a sufrir algún ataque o agresión de carácter grave. Para ello, el ASR utiliza diversos indicadores o variables relacionadas principalmente con los antecedentes judiciales de las víctimas e imputados, además de una herramienta de procesamiento de lenguaje natural que detecta palabras o analiza emociones a partir de las denuncias, comunicaciones o contactos de la víctima con el sistema procesal penal. Si bien el riesgo parece no estar determinado únicamente por estos indicadores, es probable que estas variables terminen impactando en ciertas víctimas o en ciertos imputados que no tienen una trayectoria judicial previa. Ello podría afectar, por ejemplo, a mujeres de sectores altos, respecto de las cuales existe cierta evidencia de que están menos dispuestas a denunciar por una serie de

⁹¹ GINÈS I FABRELLAS (2022), pp. 295-332.

⁹² Puede existir una cuarta forma, relacionada con los impactos de la implementación del algoritmo. Aun con un diseño cuidadoso, los algoritmos pueden causar discriminación debido al contexto en el que se implementan. Si un sistema se usa en una población sin acceso adecuado a Internet o con habilidades digitales limitadas, la discriminación surge por las brechas tecnológicas y no por el algoritmo en sí.

⁹³ BAROCAS y SELBST (2016), pp. 671-732.

factores. Si bien diversos estudios muestran que la violencia doméstica de género no impacta a todas las mujeres por igual, especialmente por la condición socioeconómica de las víctimas, existe cierta evidencia de que, al igual que sucede en todos los sectores, hay una reticencia a denunciar o reportar estos hechos por diferentes motivos.⁹⁴

4.2.2. Datos de Entrenamiento o de Retroalimentación del Algoritmo

Los datos que alimentan a los algoritmos pueden ser inexactos, insuficientes o contener sesgos de múltiple carácter. En términos coloquiales, en ciencias de la computación se conoce "*garbage in, garbage out*". Los problemas pueden originarse en los datos de entrenamiento, es decir, en las bases de datos que se utilizaron para testear el algoritmo y evaluar su rendimiento, o en los datos que retroalimentan el algoritmo en su fase de implementación. Tal como se señala en la PNIA actualizada, “[a]unque la existencia de grandes volúmenes de datos no es garantía de buenos algoritmos de IA, es necesario para el entrenamiento adecuado de los algoritmos. Además, se requiere que los datos representen adecuadamente las realidades (sean de buena calidad), sin sesgos nocivos, interoperables, reutilizables, correctamente anonimizados y agregados (si corresponde), no discriminatorios, entre otras características.”⁹⁵

En primer término, los sesgos discriminatorios pueden derivar del hecho de que las bases de datos no representan adecuadamente a todos los grupos sociales. Si las bases de datos del ASR, compuestas en su mayor parte por los antecedentes judiciales tanto de la víctima como del imputado, tienen a sobre representar ciertos grupos, es probable que el rendimiento o la precisión del algoritmo se comporte de peor manera con ciertos grupos. Por otra parte, y tal como nos comunicó uno de los funcionarios que participó en las fases iniciales de diseño del algoritmo, el objetivo es que, en el futuro próximo, el ASR incluya bases de datos más allá de aquellas que están en poder del Ministerio Público. En efecto, de incluir bases de datos como aquellas contenidas en el Registro Social de Hogares o en otros contactos de la víctima con diversos servicios públicos, tal como sucede en el VIOGEN, el ASR podría adquirir mayores niveles de precisión y demostrar un mejor rendimiento. Por ejemplo, quizás un factor de riesgo determinante sea el cobro de los beneficios sociales, la asistencia a controles de salud, o la asistencia de niños al colegio. En este escenario, las bases de datos tendrían muy buena información de personas que dependen de los servicios de asistencia social, pero no así respecto de otros grupos que tienen menos contactos con las agencias estatales. Si el “Estado de Bienestar Digital” está compuesto de bases de datos muy completas respecto de ciertos grupos, pero muy incompleta respecto de otros, es posible que ello genere una suerte de discriminación contra grupos que dependen menos de los servicios de asistencia social.⁹⁶ Además, considerando que la herramienta de PLN que incluye el ASR permite etiquetar los cuerpos de textos con nuevos factores de riesgo, es evidente que el etiquetado de esos datos puede estar contaminado por los sesgos de los propios funcionarios (*labelling bias*).⁹⁷

⁹⁴ KIBRIS y NELSON (2023), pp. 963-978.

⁹⁵ PNIA actualizada 2024

⁹⁶ VAN TOORN *et al*, pp. 507-522.

⁹⁷ BAROCAS y SELBST (2016), pp. 671-732.

De acuerdo con la información recabada, no existen antecedentes de que las bases de datos se encuentren desagregadas por criterios demográficos o de otro tipo, de modo que permitan medir o analizar la equidad estadística o aplicar otras métricas de *fairness* que son tradicionales en la evaluación o auditorías que deberían hacer los sistemas automatizados que utilizan herramientas de IA catalogadas de alto riesgo.⁹⁸ En efecto, y tal como señala la PNIA citada antes, la anonimización y agregación de datos no es lo mismo, y en algunos casos puede requerirse la desagregación de datos para que se pueda medir la proporcionalidad del impacto en diversos grupos demográficos. Uno de los funcionarios entrevistados reconoció que el ASR no incluye ciertos factores que, de acuerdo con la expertise profesional, son determinantes para hacer una evaluación del riesgo, como la nacionalidad, el estatus migratorio o la identidad de género.

“el factor de la migración es un factor evidente de la vulnerabilidad (...) nosotros hemos tratado de suplirlo porque dentro de nuestras plataformas sólo aparece la nacionalidad, este no es un dato que uno saque fácilmente como como pasa con otros datos que están súper claros, de hecho nosotros lo hemos tratado de sacar para diversos ámbitos, pero no lo sacamos fácilmente. Pero efectivamente ese (migración) es un dato que debiera estar mucho más parametrizado, así como está el parámetro de hombre o mujer que todavía está bastante binario en nuestro sistema digamos, pero esas cosas deberían estar mucho más claras para poder obtener el dato de manera mucho más fácil.” (Funcionario 2)

Otro funcionario señala algo parecido, incluyendo lo problemático que resulta que el ASR no incluya, por ejemplo, el factor de identidad de género:

“Normalmente este dato (migración) se obtiene de los informes, cuando uno hace el informe dice víctima haitiana de tantos años de edad, ahí se va levantando, pero debiéramos mejorar en ese ámbito nuestro sistema, lo hemos planteado, comunidades de víctimas lo han planteado en diversas oportunidades, el tema de la gente trans, etcétera, porque todo eso va generando factores de vulnerabilidad que son importantes de pesquisar rápido. Nosotros hemos incorporado la pregunta porque muchas veces no sabemos, y partimos de la pregunta ¿cómo le gusta a usted que la llamen? y ahí nos enteramos que la persona tiene una identidad de género distinta (funcionario 3).

De acuerdo con la nueva normativa, es plausible sostener que existe un “interés legítimo” para que el ASR pueda recolectar datos personales o incluso sensibles que permitan medir el impacto desproporcionado que puede tener el rendimiento de este algoritmo público en ciertos grupos sociales.⁹⁹

4.2.3. Aprendizaje del Algoritmo

En su etapa de implementación, los algoritmos pueden aprender correlaciones y patrones que, aunque inicialmente parecen inofensivos, resultan discriminatorios para ciertos grupos. Por

⁹⁸ HELLMAN (2020), pp. 811-866

⁹⁹ Artículo 13, letra d) de la nueva ley 19.628.

ejemplo, y tal como ha sucedido en la práctica, un algoritmo que seleccione candidatos a un puesto de trabajo puede asociar erróneamente la tenencia de perros con la idoneidad para el cargo, perjudicando a aquellos que no cumplen con esta característica, muchos de los cuales pueden pertenecer a grupos desaventajados. Lo que busca el algoritmo es la correlación, no la causalidad, y si lo primero se puede detectar de manera más eficiente a partir de vínculos puramente espurios o que no tienen una explicación aparente, entonces el sistema aprenderá a razonar de ese modo. En efecto, es posible que ciertas palabras o causas que aparentemente no están relacionadas con factores de riesgo (por ejemplo, la presencia de mascotas en el hogar, o la práctica de ciertos deportes por parte del imputado) puedan aparecer como indicadores más eficientes para gatillar alertas altas. Nuevamente, tal como señalamos antes, ello puede no parecer problemático si el interés es la protección de las víctimas. Sin embargo, si esos grupos, que aparecen en principio como normativamente irrelevantes (por ejemplo, imputados que practican fútbol), terminan viendo afectados sus derechos en mayor proporción que otros, estamos ante un caso de discriminación algorítmica.¹⁰⁰ Además, en caso de que esos grupos, a su vez, muestren cierta correlación con categorías tradicionalmente protegidas por el derecho de la antidiscriminación, se podría dar un caso de discriminación por asociación, ya sea de naturaleza directa (en caso de que la correlación sea muy intensa) o indirecta, dependiendo de las circunstancias.

Conclusiones

Con la aprobación y pronta entrada en vigencia de una nueva normativa en materia de protección de datos personales, que apunta no sólo a la protección de la privacidad y autodeterminación informativa de los titulares de datos personales, sino de todos sus derechos, es necesario que todos los órganos del Estado adapten su normativa y funcionamiento a los nuevos estándares. En este trabajo, y a través de una metodología integral, hemos descrito el modo en que funciona un algoritmo público que automatiza la evaluación del riesgo de las víctimas de violencia de género. Si bien el ASR se encuentra en estado de pilotaje, es importante partir de la base de que un algoritmo piloteado o testeado con casos reales es un algoritmo implementado, al que se le aplican todas las normas o estándares que sean pertinentes. Además, hemos destacado algunos de los puntos o notas positivas, como las diferencias que se observan en el funcionamiento del VIOGEN español y el ASR chileno: mientras el primero constituye una plataforma de acceso masivo por diversos intervinientes, el segundo parece estar controlado por personas que tienen una destacada experiencia en el tratamiento de este tipo de casos. Este último, al ser un algoritmo que meramente “sugiere” un riesgo y una determinada orientación para el diseño de las medidas de protección hacia la víctima o la solicitud de medidas cautelares para el imputado, no pretende reemplazar el juicio de los expertos. Además, a diferencia del VIOGEN, el ASR parece estar diseñado para que los fiscales, encargados de la persecución penal y de la protección de las víctimas y testigos, puedan comprender no sólo el funcionamiento general del sistema, sino la explicabilidad *ex post*. De otro modo, la utilidad del ASR podría ser cuestionada por las defensas de los imputados, por afectar el debido proceso de sus representados. El Ministerio Público, así como todos los órganos del Estado, tienen la

¹⁰⁰ WACHTER (2022), *passim*.

CODDOU MC MANUS, Alberto; RUIZ-ESQUIDE ENRÍQUEZ, Isabel; BERGERET SÁEZ, Dominique; LOYOLA OLMOS, Camila; SÁNCHEZ ANDRADE, Valentina: “IA y violencia contra las mujeres: la automatización de la evaluación del riesgo”.

obligación de adecuar tanto sus normativas internas como el funcionamiento y capacitación de sus funcionarios para enfrentar los desafíos que plantea una realidad marcada por la intermediación algorítmica de las relaciones entre el Estado y sus ciudadanos. Una vez aprobada, la nueva normativa otorgará un plazo de adecuación de dos años, que creemos es un tiempo suficiente para que la Fiscalía Nacional pueda reflexionar acerca de los temas que hemos abordado aquí.

Bibliografia citada

CODDOU MC MANUS, Alberto; RUIZ-ESQUIDE ENRÍQUEZ, Isabel; BERGERET SÁEZ, Dominique; LOYOLA OLMOS, Camila; SÁNCHEZ ANDRADE, Valentina: “IA y violencia contra las mujeres: la automatización de la evaluación del riesgo”.

- ALMOND, Louise, MCMANUS, Michelle, BRIAN, David y MERRINGTON, Daniel (2017): “Exploration of the risk factors contained within the UK’s existing domestic abuse risk assessment tool (DASH): do these risk factors have individual predictive validity regarding recidivism?” en: *Journal of Aggression, Conflict and Peace Research*, (Vol. IX) pp.58-68.
- ALON-BARKAT, Saar y BUSUIOC, Madalina (2023): “Human–AI Interactions in Public Sector Decision Making: “Automation Bias” and “Selective Adherence” to Algorithmic Advice”, en: *Journal of Public Administration Research and Theory*, (Vol. XXX, Issue 1), pp. 153-169.
- AMUSA, Lateef, BENGESAI, Annah y KHAN, Hafiz (2022). “Predicting the Vulnerability of Women to Intimate Partner Violence in South Africa: Evidence from Tree-based Machine Learning Techniques”, en: *Journal of Interpersonal Violence* (Vol. XXXVII).
- Banco Mundial (2021): Informe “Hoja de Ruta para el Diseño de una Plataforma Integrada de Gestión de Casos de Violencia contra las Mujeres para el Estado Chileno”. Disponible en: <https://documents1.worldbank.org/curated/en/755671624607466252/pdf/Hoja-de-Ruta-para-el-Disenio-de-una-Plataforma-Integrada-de-Gestion-de-Casos-de-Violencia-contra-las-Mujeres-para-el-Estado-Chileno.pdf> [visitado el 13/10/2023].
- BAROCAS, Solon y SELBST, Andrew (2016): “Big data’s disparate impact”, en: *California Law Review* (vol 104, issue 3), pp. 671-732.
- BRAVO, Erika y BARRA, Carmen (2022): “Pauta Unificada Inicial de Riesgo (PUIR) Ley N° 21.057, en PPT Capacitación PJUD. Disponible en <https://www.pjud.cl/prensa-y-comunicaciones/docs/download> [visitado el 15/11/2023].
- BUZAWA, Eve y BUZAWA, Carl (2017): *Global Responses to Domestic Violence* (Suiza, Springer).
- CARRASCO JIMÉNEZ, Edison (2018): “Reglas procedimentales sobre el fenómeno de la violencia intrafamiliar”, en: *Actualidad Penal*, (núm. 52), pp. 209-231.
- CASAS BECERRA, Lidia y VARGAS PAVEZ, Macarena (2011) : “La respuesta estatal a la violencia intrafamiliar”, en: *Revista de derecho (Valdivia)* (Vol.24, n°1), pp.133-151.
- Circuito Intersectorial de Femicidio (2021): “Informe Anual 2020”. disponible en <https://oig.cepal.org/es/documentos/informe-circuito-intersectorial-femicidio-ano-2020>. [visitado el 10/10/2023].
- Centro de Estudios y Análisis del Delito (2020): “Encuesta de violencia contra la mujer en el ámbito de violencia intrafamiliar y en otros espacios”. Disponible en: <https://cead.spd.gov.cl/centro-de-documentacion/>. [visitado el 10/09/2023].
- Centro de Estudios y Análisis del Delito (2022): “Estadísticas delictuales año 2022”. Disponible en: <https://cead.spd.gov.cl/estadisticas-delictuales/>. [visitado el 03/09/2023].
- CONTRERAS, Pablo (2020): “El derecho a la protección de datos personales y el recogimiento de la autodeterminación informativa en la Constitución chilena”, en: *Estudios Constitucionales* (Vol.18, N°2), pp. 87-120.

- Convención sobre la Eliminación de Todas las Formas de Discriminación contra la Mujer (2022): “IX informe CEDAW”. Disponible en: <https://www.exteriores.gob.es/es/PoliticaExterior/Documents/DDHH/IX%20INFORME%20CEDAW.pdf> [visitado el 12/07/2024].
- CUMMINGS, Mary (2006): “Automation and accountability in decision support system interface design”, en: *The Journal of Technology Studies* (Vol. 32, N°1), pp. 23-31.
- Escuela de Gobierno, Laboratorio de Innovación Pública (GobLab) de la Universidad Adolfo Ibáñez (2023): “Repositorio de Algoritmos Públicos. Informe anual 2023”. Disponible en https://www.algoritmospublicos.cl/static/Informes/GobLab-UAI_Informe_Repositorio_Algoritmos_Publicos_2023.pdf [visitado el 30/08/2023].
- Éticas (2022): “The external audit of the VioGén System”. Disponible en: <https://eticasfoundation.org/auditorias-externas-algoritmos/> [visitado el 30/08/2023].
- Fiscal Nacional (2021): “Oficio N.º 1032/2021. Instrucción general que imparte criterios de actuación en delitos de violencia de género y violencia intrafamiliar”. Disponible en <http://www.fiscaliadechile.cl/Fiscalia/instructivos/index.do> [visitado el 15/08/2023].
- Fiscalía Nacional (2022): “Informe Estadístico. Homicidios con perspectiva de género”. Disponible en http://www.fiscaliadechile.cl/Fiscalia/Informe_Femicidios_v2.pdf. [visitado el 10/10/2023].
- GINÈS I FABRELLAS, Anna (2022): “Sesgos discriminatorios en la automatización de decisiones en el ámbito laboral: evidencias de la práctica”, en: *Discriminación algorítmica en el ámbito laboral* (Madrid, Thompson Reuters Aranzadi), pp.295-332.
- Gobierno de Chile (2023): Comunicado de prensa “Presidente de la República Gabriel Boric Font se reúne con agrupaciones de mujeres tras la entrada en vigencia de la Ley de Reparación para Víctimas de Femicidio y sus Familias”. Disponible en: <https://prensa.presidencia.cl/comunicado.aspx?id=233851>. [visitado el 05/09/2023].
- GONZÁLEZ, José Luis (2018): “sistema de seguimiento integral en los casos de violencia de género (sistema VioGén)”, en: *Cuadernos de la guardia civil N°56*, España. Disponible en: https://gcivil.orex.es/local_repository/koha_upload/a0f8004e26b05758e7dfc9384241a959_5.pdf [visitado el 20/10/2023]
- GONZÁLEZ, José, LÓPEZ, Juan y MUÑOZ, Marina (2018): *La Valoración policial del riesgo de violencia contra la mujer pareja en España – Sistema VioGén*, disponible en: https://www.interior.gob.es/opencms/pdf/archivos-y-documentacion/documentacion-y-publicaciones/publicaciones-descargables/seguridad-ciudadana/La_valoracion_policial_riesgo_violencia_contra_mujer_pareja_126180887.pdf. [visitado el 28/08/2023].
- GONZÁLEZ GIL, Teresa (2009): “Flexibilidad y reflexividad en el arte de la investigación cualitativa”, en: *Revista Index Enferm* (Vol XIII), pp. 121-125.
- HELLMAN, Deborah (2020): "Measuring algorithmic fairness", en: *Virginia Law Review* (Vol. 106, N°4), pp. 811-866.

CODDOU MC MANUS, Alberto; RUIZ-ESQUIDE ENRÍQUEZ, Isabel; BERGERET SÁEZ, Dominique; LOYOLA OLMOS, Camila; SÁNCHEZ ANDRADE, Valentina: "IA y violencia contra las mujeres: la automatización de la evaluación del riesgo".

HOLTON, Robert y ROSS, Boyd (2021): "'Where are the people? What are they doing? Why are they doing it?'(Mindell) Situating artificial intelligence within a socio-technical framework.", en: *Journal of Sociology* (Vol. LVII), pp.179-195.

JANSEEN, Heleen (2020): "An approach for a fundamental rights impact assessment to automated decision-making", en: *International Data Privacy Law* (Volume 10, Issue 1), pp. 76-106.

KAMINSKI, Margot (2021): "The right to explanation, explained", en: SANDEEN, Sharon (Dir.), *Research Handbook on Information Law and Governance* (Edward Elgar Publishing), pp. 278-299.

KIBRIS, Arzu y NELSON, Phillip (2023): Female income generation and intimate partner violence: Evidence from a representative survey in Turkey, en: *Journal of International Development* (Volume 35, Issue 6), pp. 963-978.

Ley N.º 19.968, crea los Tribunales de Familia, publicada en *Diario Oficial* el 30 de agosto de 2024.

Ley N.º 21.378, establece monitoreo telemático en las leyes N.º 20.066 y N.º 19.968, publicada en *Diario Oficial* el 04 de octubre de 2021.

Ley N.º 21.675, estatuye medidas para prevenir, sancionar y erradicar la violencia en contra de las mujeres, en razón de su género, publicada en *Diario Oficial* el 14 de junio de 2024.

LÓPEZ DE ZUBIRÍA DÍAZ, Sandra (2024): "Discriminación algorítmica e ilusoria neutralidad. Actuales desafíos ante el fortalecimiento de tradicionales desigualdades en materia de violencia contra la mujer", en *Revista FEMERIS: Revista Multidisciplinar de Estudios de Género* (Vol. IX), pp.77-94.

Machado, Paulo, PAIS, Lucia, FELGUEIRAS, Sergio y QUARESMA, Carina (2021). "Frontline responses to high impact domestic violence in Portugal" en LOBNIKAR Branko, VOGT, Catharina y KERSTEN J. (Eds.). *Improving frontline responses to domestic violence in Europe* (Maribor, University of Maribor Press), pp. 215-239.

Ministerio de Igualdad España (2022): "Estrategia Estatal para combatir las violencias machistas 2022-2025". Disponible en: https://violenciagenero.igualdad.gob.es/planActuacion/estrategiasEstatales/combaterViolenciaMachista/estrategia_2022_2025.htm#:~:text=La%20%E2%80%9CEstrategia%20estatal%20para%20combatir,las%20violencias%20contra%20todas%20las [visitado el 20/10/2023]

NOVITZKY, Peter, JANSSEN Janine y KOKKELER, Ben (2023): "A systematic review of ethical challenges and opportunities of addressing domestic violence with AI-technologies and online tools", en *Revista Heliyon* (Vol. IX, N°6) pp. 1-15.

Organización Mundial de la Salud (2021): Comunicado de prensa conjunto "La violencia contra la mujer es omnipresente y devastadora: la sufren una de cada tres mujeres". Disponible en: <https://www.who.int/es/news/item/09-03-2021-devastatingly-pervasive-1-in-3-women-globally-experience-violence>. [visitado el 10/06/2024].

- Red Europea de Prevención del Delito (2021): “Sistema integrado sobre Violencias de Género”. Disponible en: <https://eucpn.org/document/viogen> [visitado el 25/07/2024].
- REISMAN, Dillon, SCHULTZ, Jason, CRAWFORD, Kate, WHITTAKER, Meredith (2018): "Algorithmic impact assessments: a practical Framework for Public Agency", en: *AI NOW* (9).
- ROA, Marcela. SANABRIA-MOYANO, Jesús. DINAS-HURTADO, Katherin (2023): “Herramientas de predicción de violencia basada en género y femicidio mediante la Inteligencia Artificial”, en: *Revista Jurídica Mario Alario D’Filippo* (Vol. XV), pp. 360-390.
- Satariano, A y Toll Pifarré (2024): “An Algorithm Told Police She Was Safe. Then Her Husband Killed Her” en *The New York Times*. Disponible en: <https://www.nytimes.com/interactive/2024/07/18/technology/spain-domestic-violence-viogen-algorithm.html> [visitado el 20/07/2024]
- Subsecretaría del Interior (2019): “Ordinario N.º 18.311 que informa sobre solicitud de información N°1 AB001W0010259, según lo dispuesto en el inciso segundo del artículo 10 de la Ley N.º 20.285, sobre acceso a la información pública” disponible en <https://www.interior.gob.cl/transparencia/doc/ActosSecretos/3105.pdf> [visitado el 15/08/2023].
- SVALIN, Klara, MELLGREN Caroline, TORSTENSSON Marie, LEVANDER, Sten (2018): “Police employees' violence risk assessments: The predictive validity of the B-SAFER and the significance of protective actions”, en *Revista International Journal of Law and Psychiatry* (Vol. 56) pp. 71-79.
- TURNER, Emily, BROWN, Gavin y MEDINA-ARIZA, Juanjo (2022): Predicting Domestic Abuse (Fairly) and Police Risk Assessment, en *Revista Psychosocial Intervention* (Vol. XXXI, N°3) pp 145 - 157.
- VALENCIA, Ángel (2023) en cuenta pública de la Fiscalía de Chile 2023. Disponible en: http://www.fiscaliadechile.cl/Fiscalia/quienes/cuenta_publica2023pdf.pdf
- VAN TOORN, Georgia, HENMAN, Paul, SOLDATIC, Karen (2024): "Introduction to the digital welfare state: Contestations, considerations and entanglements”, en: *Journal of Sociology* (Volume 60, Issue 3), pp. 507-522.
- VILLEGAS DÍAZ, Myrna (2017): *Violencia contra la mujer en Chile y Derechos Humanos. Informe temático 2017* (Santiago, Centro de Derechos Humanos, Facultad de Derecho Universidad de Chile).
- VIOLLIER, Pablo y FISCHER, Emma (2023): “La intervención humana como resguardo ante la toma automatizada de decisiones: implicancias éticas y jurídicas”, en: AZUAJE, Michelle (Dir.), *Introducción a la ética y el derecho de la inteligencia artificial* (Santiago, La Ley), pp. 151-169.
- WACHTER, Sandra (2022): “The theory of artificial immutability: Protecting algorithmic groups under anti-discrimination law”, en: *Tulane Law Review* (Vol. 97, N°2), pp. 1-50.

CODDOU MC MANUS, Alberto; RUIZ-ESQUIDE ENRÍQUEZ, Isabel; BERGERET SÁEZ, Dominique; LOYOLA OLMOS, Camila; SÁNCHEZ ANDRADE, Valentina: “IA y violencia contra las mujeres: la automatización de la evaluación del riesgo”.

ZURITA, Jorge (2014): “Violencia contra la mujer: Marco histórico evolutivo y predicción del nivel de riesgo” en Tesis doctoral, Universidad Autónoma de Madrid. Disponible en <https://produccioncientifica.ucm.es/documentos/5d39994d2999520684440fc1>

Jurisprudencia citada

STS* (Sala de los Contencioso-Administrativo) de 30 de septiembre de 2022 (SAN 2350/2020).