GUÍA

Formulación ética de proyectos

DE CIENCIA DE DATOS

BID | LAB
UNIVERSIDAD ADOLFO IBÁÑEZ
gob digital CHILE

Guía
Formulación ética de proyectos de ciencia de datos

Agosto 2022

Citación sugerida
Formulación ética de proyectos de ciencia de datos (2022), División de Gobierno Digital y Universidad Adolfo Ibáñez.

El desarrollo de esta guía ha sido posible gracias al proyecto Algoritmos Éticos, Responsables y Transparentes, que cuenta con apoyo de BID Lab —el laboratorio de innovación del Banco Interamericano de Desarrollo— y un equipo mixto de profesionales de la División de Gobierno Digital (DGD) y la Universidad Adolfo Ibáñez (UAI).

Equipo UAI
Gabriela Denis Jara
Romina Garrido Iglesias
María Paz Hermosilla Cornejo
Reinel Tabares Soto.

Equipo DGD
Jaime Astorquiza Lumsden
Braulio Neira Paredes
Kareen Schramm López
Christian Sifaqui Merczak.

Diseño editorial
Estudio Real
somosreal.cl

BID LAB UNIVERSIDAD ADOLFO IBÁÑEZ gob digital CHILE

Contenido

1 Acerca de esta guía 4

1.1 Objetivo de esta guía 5
1.2 ¿A quién va dirigida? 5
1.3 Proyecto Algoritmos Éticos, Responsables y Transparentes 7

2 Introducción 8

3 Consideraciones legales y éticas 11

3.1 Uso y protección de datos personales 12
3.1.1 Normativa chilena 12
3.1.2 Consideraciones éticas: más allá de la ley 19
3.2 Transparencia y rendición de cuentas (accountability) 25
3.2.1 Normativa chilena 26
3.2.2 Consideraciones éticas: más allá de la ley 28
3.3 Sesgos y discriminación 31
3.3.1 Normativa chilena 33
3.3.2 Consideraciones éticas: más allá de la ley 35
3.4 Otras consideraciones éticas 37

4 Casos de uso y riesgos legales y éticos 41

4.1 Sistemas de reconocimiento y de detección de eventos 43
4.2 Predicción 46
4.3 Personalización 49
4.4 Soporte de interacción 50
4.5 Optimización 52
4.6 Razonamiento con estructuras de conocimiento 54

5 Conclusiones 56

Referencias 58
Anexo 1: Glosario 59

01

Acerca de esta guía

1.1 Objetivo de esta guía

1.2 ¿A quién va dirigida?

1.3 Proyecto Algoritmos Éticos, Responsables y Transparentes

1. Acerca de esta guía

1.1 Objetivo de esta guía

El objetivo de esta guía es proveer al sector público en Chile una base de conocimiento mínimo para la formulación ética de proyectos en ciencia de datos conociendo los riesgos a la hora de formular y entrenar modelos.

1.2 ¿A quién va dirigida?

Esta guía va dirigida a funcionarios e instituciones públicas que tienen planificado desarrollar proyectos tecnológicos que involucran el uso intensivo y el análisis de datos para mejorar su gestión o entrega de servicios a las personas. Específicamente, de sistemas de toma o soporte de decisión donde existen potenciales riesgos éticos y legales, como sesgos y protección de datos personales. Entenderemos sistemas de toma y soporte de decisión como (González, Ortiz & Sánchez, 2020):

Sistemas de toma de decisión

Aquellos sistemas en los cuales la decisión es completamente automática, donde las decisiones finales se toman sin intervención humana.

Sistemas de soporte de decisión

Aquellos sistemas que son utilizados para informar la acción a realizar por una persona.

Por lo tanto, esta guía se enfocará en modelos algorítmicos que ayudarán en el proceso de toma de decisiones al interior de las instituciones y que impactarán su accionar. Esta guía no está enfocada en licencias de software o hardware.

Por ejemplo, esta guía está pensada para proyectos de los siguientes tipos:

1. Sistemas de decisión automáticos que realizan **tratamiento de datos personales** y/o sensibles con consideraciones sobre la privacidad.
2. Sistemas de decisión automáticos que tienen un **efecto significativo en la vida de las personas**. Por ejemplo: suspensión o cancelación de un contrato, derecho o denegación de una prestación concedida por Ley, denegación de ciudadanía o de admisión a un país, elegibilidad a un crédito, acceso a servicios sanitarios o sociales, decisiones con respecto a la contratación, promoción o despido de un empleo, acceso a la educación, entre otros.
3. Sistemas de decisión automáticos destinados a la **fiscalización, detección de fraudes**, cursar multas u otras tareas afines relacionadas con el ejercicio de las potestades sancionatorias del Estado.
4. Sistemas de decisión automáticos de **predicción sobre el comportamiento** de las personas.
5. **Sistemas de agrupación** de personas con características predeterminadas.
6. Sistemas de decisión automáticos en el **área de salud**, detección de enfermedades, apoyo al diagnóstico médico o la gestión de la red de salud asistencial.
7. Sistemas de decisión automáticos de **resolución de conflictos** o de apoyo a la decisión humana.
8. Sistemas de decisión automáticos que buscan **mejorar la gestión** interna de instituciones.
9. Sistemas de **reconocimiento de patrones** (imágenes, cosas, sentimientos, etc.)
10. Sistemas de decisión automáticos que afectan a p**ersonas o grupos vulnerables** de la sociedad y/o que reciben protección social del Estado.[1]

1. Esta lista es una sugerencia, pudiendo existir otros tipos de sistemas no presentes en esta.

1.3

Proyecto Algoritmos Éticos, Responsables y Transparentes

El desarrollo de esta guía se enmarca en el proyecto Algoritmos Éticos, Responsables y Transparentes[2], que es ejecutado por la Universidad Adolfo Ibáñez (UAI) con el apoyo de BID Lab, el laboratorio de innovación del Grupo BID, y es parte de su iniciativa fAIrLAC, la que promueve la aplicación ética y responsable de la inteligencia artificial en la región, ayudando al sector público y privado a mejorar la provisión de servicios sociales y el desarrollo de empresas de impacto social.

Este es un proyecto inédito en Chile, que busca instalar capacidades y estándares para incorporar consideraciones éticas tanto en la compra y utilización de inteligencia artificial y algoritmos de decisión automatizada en agencias estatales, como en la formulación y desarrollo de estas soluciones por parte de los proveedores tecnológicos, creando con ello oportunidades de mercado para el sector.

Esta guía es parte de la línea de trabajo de formación de capacidades del proyecto, cuyo trabajo también incluye la comprobación empírica de herramientas desarrolladas por la iniciativa fAIr LAC, el Foro Económico Mundial y otras entidades, a través de pilotos con distintas instituciones públicas.

2 https://goblab.uai.cl/algoritmos-eticos/

02

Introducción

2. Introducción

Los proyectos de tecnología cada vez se hacen más cotidianos y es natural que el siguiente paso sea cómo implementar proyectos de esa naturaleza al interior de las instituciones, buscando solucionar problemas reales y significativos, formulando proyectos que tengan un impacto directo en el bienestar de la ciudadanía o de colaboradores internos.

En 2019 se aprobó la Ley 21.180 sobre transformación digital del Estado, que busca hacerlo más transparente, moderno y capaz de entregar mejores servicios a la ciudadanía. Esta ley declara la necesidad de contar con documentos digitales, compartir información entre instituciones y digitalizar en lo posible los trámites administrativos. La entrada en vigencia de esta ley es, por lo tanto, una oportunidad para que las instituciones puedan utilizar todos los datos disponibles en beneficio de la ciudadanía.

Los proyectos de inteligencia artificial (IA) y de ciencia de datos, como cualquier otro, deben estar siempre circunscritos a las leyes vigentes en el país pero, más aún, tomar en cuenta los riesgos éticos que existen al desarrollar un proyecto de esas características. Los proyectos de datos y las decisiones algorítmicas se dan en un contexto social y pueden tener efectos no deseados para la población. Estos proyectos, al involucrar todo un sistema de decisión, deben tener en cuenta conceptos como la proporcionalidad (**¿es realmente necesario un algoritmo?**) y la explicabilidad (**¿qué factores toma en cuenta mi algoritmo?**, **¿son suficientes?**, **¿podemos explicar y es entendible el proceso para llegar al resultado?**) de los algoritmos, así como también disponer de un canal para rectificar potenciales efectos adversos.

¿Es realmente necesario un algoritmo?

¿Qué factores toma en cuenta mi algoritmo? ¿Son suficientes?

¿Podemos explicar y es entendible el proceso para llegar al resultado?

Una buena práctica es considerar las limitaciones legales y éticas desde la formulación del proyecto ya que, si no se hace, es probable que dichos riesgos se hagan presentes en la etapa de ejecución cuando ya poco puede hacerse para mitigarlos. Tener claridad sobre potenciales riesgos éticos y legales desde el primer momento ayuda a crear sistemas de respuesta y de mitigación de los mismos, así como también de comunicación de beneficios de la herramienta, lo cual es favorable para cualquier proyecto de ciencia de datos.
Dado que el equipo técnico (interno o externo) que desarrollará la tecnología puede no tener conocimiento pleno de todos los lineamientos legales que impactarán al proyecto, es importante que

quien esté a cargo del proyecto fomente desde una etapa temprana -desde su formulación- la comunicación entre el equipo técnico y el equipo legal de la institución.

En esta guía se entenderá por formulación todo el proceso previo a la obtención de financiamiento para la implementación de un proyecto o a la obtención de apoyo para implementarlo con recursos institucionales propios. Es decir, comprende todo el proceso de diseño previo a la implementación de la herramienta o modelo.

Esta guía entregará los principales lineamientos éticos y legales que se deben tener en cuenta cuando se está formulando un proyecto de tecnología. En particular, se enfocará únicamente en aquellos proyectos que sean sistemas de soporte o de toma de decisión automáticos, que buscan cambiar el accionar actual de la institución y la forma en que se da respuesta actualmente al problema que se quiere solucionar.

Este documento se divide en dos grandes partes. En la primera se detallarán algunos de los aspectos legales y éticos más relevantes a tener a la vista en proyectos de esta naturaleza (Capítulo 3), y en la segunda parte se ejemplificará aquellos más relevantes según tipo de tarea a través de casos de uso chilenos (Capítulo 4). Se utilizará para ello la clasificación de la OECD (OECD, 2022) de sistemas de IA.

Los casos de uso en Chile fueron obtenidos del repositorio de Algoritmos Públicos del GobLab UAI[3] y del proyecto realizado por el GobLab UAI con el Consejo para la Transparencia para el diagnóstico sobre transparencia algorítmica en sistemas de decisión automatizadas y semi automatizadas en el Estado de Chile[4].

3 https://www.algoritmospublicos.cl/
4 https://goblab.uai.cl/transparencia-algoritmica-en-el-sector-publico/

03

Consideraciones legales y éticas

3.1 Uso y protección de datos personales

3.2 Transparencia y rendición de cuentas (accountability)

3.3 Sesgos y discriminación

3.4 Otras consideraciones éticas

3. Consideraciones legales y éticas

3.1 Uso y protección de datos personales

En algunos casos, los modelos de ciencia de datos utilizarán datos personales a los cuales las instituciones pueden tener acceso por sus propias competencias o por consentimiento expreso. La utilización de datos personales tiene implicancias específicas, en cuanto deberá apegarse a las leyes chilenas y tratados internacionales de Derechos Humanos suscritos por Chile, como la Declaración Universal de Derechos Humanos y el Pacto de San José, los cuales contemplan el derecho a la privacidad y al debido proceso, entre otros.

3.1.1 Normativa chilena

Leyes relevantes sobre uso y protección de datos personales:

En Chile la protección de datos personales está consagrada como un derecho en la Constitución[8], y el tratamiento y condiciones de estos datos se condiciona por la Ley N° 19.628 sobre protección de la vida privada.

5 https://www.bcn.cl/leychile/navegar?idNorma=242302&idParte=8563487
6 https://www.bcn.cl/leychile/navegar?idLey=19628
7 https://www.bcn.cl/leychile/navegar?idLey=20575
8 Numeral 4°, artículo 19 de la Constitución Política de la República https://www.bcn.cl/leychile/navegar?idNorma=242302&idParte=8563487

La primera consideración a tener en cuenta por cualquier persona que busca implementar algún sistema de toma o de soporte de decisión, es si utilizará datos personales y/o datos sensibles. En la etapa de formulación deben identificarse:

A	B	C	D
Si se trata de este tipo de datos.	Si se cuenta con las competencias legales para tratar esos datos.	Si no se cuentan con las competencias, cómo se obtendrán los datos.	Determinado el tipo de dato y los mecanismos de obtención, deberán identificarse cuáles deberían ser los principales resguardos a adoptarse durante la ejecución del proyecto.

A Identificar si se trata de datos personales:

Esta ley define los datos personales como aquellos «*relativos a cualquier información concerniente a personas naturales identificadas o identificables*». Por lo tanto, son ejemplos de datos personales nombre, RUT, domicilio, número telefónico, entre otros. En esta definición cabe destacar el uso de la palabra «identificable» en cuanto, al publicar datos personales, es importante no sólo evitar la identificación directa de la persona, sino también, evitar la identificación de la persona de manera indirecta dadas características personales que pueden llevar a ello. Por ejemplo, si solamente una persona en la base de datos posee una característica, y dicha característica es conocida, ese dato no se podrá presentar de manera desagregada ya que permitiría la identificación de la persona.

La identificabilidad ha sido un concepto delimitado en Europa por las autoridades reguladoras de protección de datos[9]. Cabe señalar que Chile carece de una, por lo tanto, se debe tener en consideración lo siguiente:

⇨ La persona física es «identificable» cuando, y aunque no se la haya identificado de manera expresa, sea posible hacerlo.

9 Documento WP 136. Dictamen 4/2007 sobre el concepto de datos personales. Disponible en https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_es.pdf

- La identificación se logra normalmente a través de datos concretos que se pueden denominar como «identificadores» y que tienen una relación privilegiada y muy cercana con una determinada persona. Así, una persona puede ser identificada directamente por su nombre y apellidos o indirectamente por un número de teléfono, la matrícula de un auto, un número de seguridad social, un número de pasaporte o por una combinación de criterios significativos (edad, empleo, domicilio, etc.) que hagan posible su identificación al estrecharse el grupo al que pertenece. Estos identificadores pueden ser datos concretos que tienen una relación cercana con determinada persona, como lo serían aquellos relativos a su apariencia física, a su profesión, etc.

- El identificador permite aislar a una persona dentro del conjunto de una población determinada. Cuando se habla indirectamente de identificadas o identificables se refiere, en general, al fenómeno de las «combinaciones únicas», sean éstas pequeñas o grandes.

- Los identificadores disponibles pueden no permitir singularizar a una persona determinada, sin embargo, ésta aún puede ser «identificable» cuando esta información se combina con otros datos, es decir, se trata de datos que necesitan el concurso de otros datos que contribuyan a identificar a su titular. En consecuencia, las asociaciones posibles y razonables de datos permitirán establecer la identidad de su titular y por ende será de aplicación la normativa de protección de datos.

Para evaluar la identificabilidad en cada caso, se deben apreciar factores objetivos como costos, tiempo necesario para la identificación, la tecnología disponible y los avances que pueda experimentar ésta al momento de producirse el tratamiento. Esto porque podría darse el supuesto de que los datos que se estén manejando hoy y que no estén en la categoría de dato personal puedan pasar a catalogarse como tal, por estar conservados por un tiempo mayor donde, dado el avance de las tecnologías, permita identificar esta información de una persona en particular.

Para que se defina un dato como personal no es necesario, entonces, una coincidencia plena entre el dato y una persona concreta, sino que además es necesario que la identificación pueda efectuarse sin esfuerzos desproporcionados.

Por su parte, los datos personales sensibles se definen como «*aquellos datos personales que se refieren a las características físicas o morales de las personas o a hechos o circunstancias de su vida privada o intimidades, tales como los hábitos personales, el origen racial, las ideologías y opiniones políticas, las creencias o convicciones religiosas, los estados de salud físicos o psíquicos y la vida sexual*»[10]. Los datos personales sensibles tienen un tratamiento distinto y más estricto que los datos personales; no se podrán utilizar datos sensibles a menos que la ley lo autorice, exista consentimiento del titular o que sean necesarios para la entrega de un tratamiento o beneficio de salud de conformidad a la normativa que regula el tratamiento de esos datos.

Datos personales	Datos personales sensibles
• Relativos a cualquier información concerniente a personas naturales, **identificadas o identificables**. • Por ejemplo, son datos personales: información relativa a la movilidad, información biométrica, información de conexiones de red, información georreferenciada, RUT, número de pasaporte, número telefónico, IMEI, entre otros. • Los datos personales pueden ser solo personales o datos personales sensibles. • Las personas jurídicas no poseen datos personales.	• La ley chilena los define como aquellos que se refieren a las características físicas o morales de las personas o a hechos o circunstancias de su **vida privada o intimidad**, tales como los hábitos personales, el origen racial, las ideologías y opiniones políticas, las creencias o convicciones religiosas, los estados de salud físicos o psíquicos y la vida sexual, etc. • Por ejemplo, son datos personales sensibles la pertenencia a una etnia, estado socioeconómico, datos que informen una opción política, datos relativos a orientación sexual, entre otros. • En otros países se denominan datos especialmente protegidos.

10 https://www.bcn.cl/leychile/navegar?idNorma=141599

B Identificar si se cuenta con las competencias para tratar los datos personales.

En el caso de las entidades públicas, para utilizar los datos personales no será necesario el consentimiento de los titulares de datos cuando éstas cuenten con las competencias para tratar los datos y lo hagan con sujeción a las reglas de la ley. El titular de los datos personales es la persona natural a la que se refieren los datos de carácter personal.

La entidad puede tener o no acceso actual a los datos que necesita en el proyecto, siendo el primer punto de verificación cómo puede obtenerlos. Si los datos están su organización, es porque tiene las competencias para tenerlos y custodiarlos. Si los datos están en otra entidad deberá celebrarse un acuerdo de colaboración. El acuerdo de colaboración no reemplaza las competencias legales, y es un documento que tiene por objeto regular las condiciones técnicas del traspaso de datos entre entidades que pueden tratarlos y establecer resguardos.

Podemos encontrar un ejemplo de competencia expresa para tratar datos personales en el DFL 1/2005 del Ministerio de Salud, que entrega al MINSAL la facultad para: «*Tratar datos con fines estadísticos y mantener registros o bancos de datos respecto de las materias de su competencia. Tratar datos personales o sensibles con el fin de proteger la salud de la población o para la determinación y otorgamiento de beneficios de salud. Para los efectos previstos en este número, podrá requerir de las personas naturales o jurídicas, públicas o privadas, la información que fuere necesaria. Todo ello conforme a las normas de la Ley N° 19.628 y sobre secreto profesional.*»

C Si no se cuenta con las competencias, definir cómo se obtendrán los datos.

Si la entidad no tiene las competencias legales para tratar esos datos deberá obtener el consentimiento del titular y sujetarse a las disposiciones de la ley, a menos que los datos provengan de una fuente de acceso público. El consentimiento o la ley son requisitos habilitantes, es decir, es el primer paso a verificar para utilizar los datos, debiendo luego aplicarse todas las reglas contenidas en la ley. La Ley N°19.628 dispone qué requisitos debe cumplir el consentimiento, los que se deben tener presente en la etapa de recopilación de datos.

D Identificar cuáles deberían ser los principales resguardos a adoptarse durante la ejecución del proyecto.

Luego, respecto de las operaciones de tratamiento de datos personales, durante la ejecución del proyecto que se formula deberán observarse las reglas contenidas en la ley, como la finalidad, confidencialidad, debida diligencia, conservación o la calidad.

En esta misma línea, la ley establece claramente que las personas a cargo de las bases de datos, en instituciones tanto públicas como privadas, deberán guardar secreto de estas cuando hayan sido recopilados por fuentes no accesibles al público. En caso de que sea un tercero el que desarrolle el modelo, será necesario un acuerdo por escrito con cláusulas de confidencialidad y las instrucciones precisas del mandato de acceso a datos, si fuese el caso, para permitir la entrega de los datos relevantes para el entrenamiento del mismo. Esto debe quedar explícito en el contrato con el proveedor, con el fin de velar por un apropiado uso de los datos.

Además, el tratamiento de los datos personales solamente podrá realizarse para los fines que fueron recolectados, a menos que se trate de datos accesibles al público, por lo cual las instituciones deben tener especial cuidado con el uso de datos personales, disponiendo también de medidas de seguridad adecuadas en su almacenamiento. La ley señala que los responsables del tratamiento deben cuidar de estos con la debida diligencia haciéndose responsables de los daños.

Además, la ley entrega diversos derechos a los titulares de datos, como el acceso, la rectificación, la eliminación y la oposición. En esta etapa el formulador del proyecto deberá identificar posibles mecanismos para el ejercicio de estos derechos, cuando correspondan. Por ejemplo, la rectificación se podrá ejercer siempre y cuando no entorpezca acciones fiscalizadoras, no afecte el derecho de reserva, no afecte la seguridad nacional y que no sean datos almacenados por mandato legal.

Por lo tanto, al utilizar datos personales y/o datos personales sensibles, el tratamiento de los mismos debe cumplir, a lo menos, con los estándares dispuestos en la Ley N° 19.628 y consultar la ley particular que puede regular alguna materia específica, cuidando siempre la privacidad de las personas, definiendo claramente su tratamiento y seguridad de almacenamiento.

La protección de datos ha sido consagrada de manera explícita en el numeral 4 del artículo 19 de la Constitución Política de la República, por lo tanto, cualquier herramienta que utilice datos personales y/o afecte la vida privada, deberá siempre cumplir con lo dispuesto en la ley y apegarse a la Constitución.

Además, más allá de lo que establece la ley general de protección de datos, siempre debe complementarse con la normativa especial que regule el ámbito de tratamiento que rige a la institución. Se deben considerar, entre otras: en el ámbito de la salud, la Ley N° 20.584 que regula los derechos y deberes que tienen las personas en relación con acciones vinculadas a su atención en salud[11]; en educación, el DFL 2 que fija texto refundido, coordinado y sistematizado de la Ley N° 20.370 con las normas no derogadas del decreto con fuerza de ley N°1, de 2005[12]; en servicios sociales, la Ley N° 21.322 que establece el sistema red integral de protección social[13]; en lo tributario, la Ley N° 21.210 que moderniza la legislación tributaria[14]. En particular, respecto del secreto tributario, la Ley N° 21.210 establece que aquellos funcionarios que accedan a información contenida en expedientes electrónicos deberán regirse por la Ley N° 19.628 de protección de la vida privada, «así como con las demás leyes que establezcan la reserva o secreto de las actuaciones o antecedentes que obren en los expedientes electrónicos»[15].

Por ejemplo, podemos mencionar a la protección del secreto estadístico como un ámbito especial del tratamiento de datos personales cuando los datos generan información estadística. La Ley N° 17.374[16] que aprobara la ley orgánica dirección estadística y censos y crea el Instituto Nacional de Estadísticas, INE, indica que esta entidad, los organismos fiscales, semifiscales y empresas del Estado, y cada uno de sus respectivos funcionarios, no podrán divulgar los hechos que se refieren a personas o entidades determinadas de que hayan tomado conocimiento en el desempeño de sus actividades estadísticas y que el estricto cumplimiento de estas reservas constituye el secreto estadístico. Su infracción por parte de cualquier persona sujeta a esta obligación hará incurrir en el delito previsto y penado por el artículo 247, del Código Penal[17], debiendo en todo caso aplicarse pena corporal. Este secreto implica que los datos estadísticos no pueden ser publicados o difundidos con referencia expresa a las personas o entidades a quienes directa o indirectamente se refieran, si mediare prohibición del o los afectados.
El secreto estadístico constituye el eje de la actividad estadística, la preservación de este secreto se traduce en la confianza de las fuentes de información mediante el resguardo del anonimato. Cabe señalar que esta obligación se extiende también, según el artículo

11 https://www.bcn.cl/leychile/navegar?idNorma=1039348
12 https://www.bcn.cl/leychile/navegar?idNorma=1014974
13 https://www.bcn.cl/leychile/navegar?idNorma=1158583
14 https://www.bcn.cl/leychile/navegar?idNorma=1142667
15 https://www.bcn.cl/leychile/navegar?idNorma=141599
16 https://www.bcn.cl/leychile/navegar?idLey=17374
17 https://www.bcn.cl/leychile/navegar?idNorma=1984&idParte=9672488

29 de la Ley N° 17.374 que crea el INE, a los organismos fiscales, semifiscales, empresas del Estado y a cada uno de sus respectivos funcionarios. La protección de la confidencialidad obliga incluso a no publicar informaciones desagregadas para evitar la posible identificación del informante.

Existen otros conceptos relevantes a considerar sobre el uso y tratamiento de datos personales. Una de las obligaciones relevantes para las entidades públicas es la inscripción de la base de datos en el Servicio de Registro Civil e Identificación, conforme al artículo 22 de la Ley N° 19.628 y el procedimiento fijado al efecto[18].

En la siguiente sección se ahondará en las consideraciones éticas.

3.1.2
Consideraciones éticas: más allá de la ley

Desde 2017, se encuentra en trámite en el Congreso una reforma a la Ley N° 19.628 sobre protección de la vida privada que busca modernizar la legislación actual para alcanzar estándares internacionales en esta materia[19]. Esta reforma incluye diversos cambios a la actual normativa, entre los cuales se pueden destacar:

- Establece nuevas bases de licitud para el tratamiento de datos personales sin necesidad de contar con el consentimiento de titulares.

- Fortalece diversos principios, entre ellos, la finalidad. Con la reforma la finalidad deberá ser específica, explícita y lícita, es decir, los datos no podrán ser tratados con fines distintos a los informados en su recolección.

- Incorpora el derecho de portabilidad de los datos personales y a la impugnación de las decisiones automatizadas, con que las personas tendrán derecho a no ser objeto de una decisión basada únicamente en el tratamiento automatizado, así como también el titular podrá impugnar y solicitar información respecto a los criterios de valoración utilizados. Este punto guarda relación con el siguiente tema de «Transparencia y rendición de cuentas».

- Establece una autoridad de protección de datos que, desde octubre de 2021, correspondería a una agencia dependiente del Ministerio de Economía para velar por el cumplimiento de lo dispuesto en la ley.

18 DS 779/2000 Justicia. https://www.bcn.cl/Leychile/navegar?idNorma=177681

19 Boletín N° 11.092-07

Es razonable esperar que en un plazo no muy lejano este proyecto se convierta en ley, estableciendo así el estándar de tratamiento de datos personales que deberán cumplir todas aquellas personas que requieran trabajar con ellos en el entrenamiento de modelos de toma o de soporte de decisión.

En otro orden de cosas, el Consejo para la Transparencia, como entidad que entre sus facultades puede velar por el cumplimiento de la Ley N° 19.628, también ha establecido recomendaciones para la protección de datos personales por organismos públicos, específicamente, elevando los estándares de políticas de seguridad de datos personales para mitigar riesgos y amenazas de confidencialidad, integridad y disponibilidad de la información, entre otras materias[20].

Con respecto al uso y tratamiento de datos personales, es relevante también plantear los desafíos de privacidad, específicamente (AEPD, 2021):

- **Privacidad desde el diseño**: Se refiere a integrar las garantías de la protección de los datos desde las primeras etapas del desarrollo de un sistema y/o producto y luego en el tratamiento mismo, entregar la mayor privacidad desde etapas tempranas del tratamiento de datos, e incluso pensando en las medidas de protección antes que el tratamiento se inicie, integrando los requisitos de la protección en el diseño del sistema.

- **Privacidad por defecto**: Este concepto complementa al anterior y actualmente las guías de protección de datos se refieren a ambos requisitos casi siempre de manera conjunta. La protección por defecto mira a las operaciones del tratamiento de datos y se relaciona con garantizar impactos mínimos al tratamiento de datos, ya sea acotando su recopilación, acceso o tiempo de almacenamiento. En la práctica, se trata de utilizar solamente aquellos datos necesarios para el fin dispuesto y definidos en la etapa de diseño. Es decir, generar un impacto mínimo y lo menos intrusivo posible en los derechos de las personas.

Es deseable, por tanto, que quienes quieren implementar proyectos de toma o soporte de decisión tengan presentes los conceptos de privacidad desde el diseño y por defecto para lograr un adecuado tratamiento de datos personales, protegiendo siempre la privacidad de las personas. En ambos casos se trata de enfocar el tratamiento de datos a la gestión del riesgo y de responsabilidad

20 Resolución Exenta N°304. Consejo para la Transparencia. https://www.consejotransparencia.cl/wp-content/uploads/estudios/2020/12/N%C2%B0304-Aprueba-el-texto-actualizado-y-refundido-de-las-recomendaciones-del-CPLT-sobre-Proteccio%CC%81n-de-Datos-Personales.pdf

de forma proactiva, mediante el uso de estrategias que incorporen la protección de la privacidad en todo el ciclo de vida del uso de los datos. La privacidad por diseño se refiere a la protección de datos personales desde el primer momento del proyecto (formulación) como, por ejemplo, anonimizando los datos; la privacidad por defecto, por su parte, tendrá probablemente mayor relevancia en la etapa de implementación del modelo, resguardando por ejemplo que no exista vulneración informática. Dado que la privacidad por defecto es relevante una vez el proyecto se implementa, escapa del alcance de esta guía.

Uso y protección de datos personales:

Preguntas claves

- **¿Estás trabajando con datos personales y/o sensibles identificables a nivel individual? ¿Cuáles? Por ejemplo: nombre (dato personal), información socioeconómica (dato personal sensible).**
- **¿Has identificado la justificación o base legal para trabajar con esos datos? ¿Está dentro de las facultades de la institución el manejo de los datos?**
- **¿Has identificado las regulaciones que podrían impactar en el proyecto? ¿Existen regulaciones específicas que afectan el proyecto, como por ejemplo en el área de salud?**
- **¿Serán necesarios mecanismos para garantizar la calidad de los datos personales, como por ejemplo mecanismos de acceso, eliminación o rectificación? ¿Cuáles?**

Uso y protección de datos personales:

Buenas prácticas

Considerar alguna metodología de **clasificación de datos** que le permita al formulador establecer en el diseño qué niveles de protección serán los necesarios cuando el proyecto comience a ejecutarse, generando un catálogo de los datos que serán objeto del tratamiento. Para estos fines, se puede utilizar la metodología presente en el documento «**Clasificación de Datos OEA**»[21]. Si bien forma parte de una obligación legal, el registro de la base de datos mandatado en la Ley N° 19.628 puede servir de base para esta de clasificación.

De acuerdo a las categorías identificadas y a los riesgos y tipos de datos tratados, decidir si será necesario realizar una **evaluación de impacto en protección de datos**. Cabe hacer presente que este no es el momento para llevar a cabo este proceso, sino para definir si será necesario o no.

Una evaluación de impacto en protección de datos es un proceso concebido para describir el tratamiento, evaluar su necesidad y proporcionalidad, y ayudar a gestionar los riesgos para los derechos y libertades de las personas físicas derivados del tratamiento de datos personales evaluándose, determinando las medidas para abordarlos. La Agencia Española de Protección de Datos publicó el documento «**Gestión del riesgo y evaluación de impacto en tratamientos de datos personales**», el cual entrega lineamientos y actividades para realizar esta evaluación de impacto.[22]

En Chile este no es un instrumento obligatorio, sin embargo, resulta una herramienta novedosa y práctica para abordar de manera eficiente los riesgos que pudieran producirse por la implementación de estos sistemas e incorporar de manera responsable la protección de datos. Y no solo eso, también puede ayudar a obtener licencia social, un componente clave en proyectos de tecnologías más disruptivas. El proyecto de ley considera que los responsables del tratamiento deberán abordar los riesgos de manera previa y la ley actual consagra la debida diligencia como un principio en esta materia.

21 https://www.oas.org/es/sms/cicte/docs/ESP-Clasificacion-de-Datos.pdf. También se puede consultar el Dictamen 4/2007 sobre el concepto de datos personales https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_es.pdf

22 Guía de Gestión de riesgo y evaluación de impacto en tratamientos de datos personales. https://www.aepd.es/es/documento/gestion-riesgo-y-evaluacion-impacto-en-tratamientos-datos-personales.pdf

Buenas prácticas

En Europa existen algunos listados orientadores de cuándo sería necesario ejecutar esta evaluación, puesto que muchas veces no resulta económicamente factible para las administraciones públicas proteger todos los datos al mismo nivel. Los niveles más altos de protección de datos acarrean costos adicionales y tienen el potencial de generar mayores gastos de lo que se merecen los datos.

Un listado de las actividades en que sería necesaria una evaluación de impacto en protección de datos que puede revisarse, a modo de orientación, es el elaborado por la Agencia Española de Protección de Datos[23], que incluye actividades de perfilamiento o evaluación sistemática y exhaustiva de los aspectos personales de una persona, tratamiento de datos a gran escala con datos sensibles -investigación con datos masivos, observación o monitoreo de zonas públicas-, y varias metodologías para este proceso si se decide realizar.[24]

Considerar elementos de la **privacidad por diseño**, que sean aplicables a esta etapa de formulación:

a. Si es necesario recopilar datos, incorporar metodologías de proporcionalidad y minimización:

   - Recoger solo datos personales que se tratarán, limitando la cantidad de datos recogidos.
   - Definir el perímetro del tratamiento.
   - Limitar el período de retención o almacenamiento.
   - Limitar el número de personas con acceso a los mismos.
   - Recogerlos solo cuando se vayan a tratar.

b. Si es necesario acceder a datos de fuentes externas, identificar si serán necesarios convenios de colaboración y los elementos mínimos a considerar en ellos.

c. Evaluar la necesidad de usar la información de manera identificada, y en caso contrario adoptar un **enfoque de preservación de la privacidad** adoptando soluciones técnicas que transforman un conjunto de datos que contiene datos personales en uno anónimo. La anonimización y la seudonimización son procedimientos claves en la gestión de datos, debiendo evaluar cuál es la técnica más adecuada al proyecto formulado.

23 https://www.aepd.es/es/documento/listas-dpia-es-35-4.pdf

24 Puede consultarse el caso de UK (https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments/) , España, Argentina (https://www.argentina.gob.ar/sites/default/files/guia_final.pdf) y Uruguay (https://www.gub.uy/unidad-reguladora-control-datos-personales/comunicacion/publicaciones/guia-evaluacion-impacto-proteccion-datos) .

Buenas prácticas

d. Una forma de integrar la privacidad en el diseño es la **ingeniería de la privacidad**, que no es otra cosa que traducir los requisitos legales que establecen derechos y libertades de los ciudadanos en guías y requisitos de los sistemas[25] o la implementación de **tecnologías de mejora de la privacidad (PET)**, que cubren la gama más amplia de tecnologías diseñadas para respaldar la privacidad y la protección de datos.[26] La elección de qué tecnología de privacidad es la adecuada de implementar dependerá -por cierto- de los objetivos de privacidad que el proyecto deba implementar, de la normativa, del uso efectivo de los datos y su utilización. Las PET permiten el análisis de datos, su uso compartido, confiable y preservando la privacidad.[27]

Crear una política de gobernanza de datos que contemple procesos estandarizados y mantenga una comunicación dentro de la institución sobre el cómo y para qué se utilizan los datos, junto con las actualizaciones de seguridad y privacidad que sean necesarias.

25 https://en.wikipedia.org/wiki/Privacy_engineering y https://www.mitre.org/publications/systems-engineering-guide/enterprise-engineering/engineering-informationintensive-enterprises/privacy-systems-engineering

26 https://www.enisa.europa.eu/topics/data-protection/privacy-enhancing-technologies y https://royalsociety.org/topics-policy/projects/privacy-enhancing-technologies/

27 https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/privacy-enhancing-technologies-report.pdf

3.2

Transparencia y rendición de cuentas (accountability)

Los algoritmos son percibidos muchas veces como cajas negras donde «entran datos y se obtiene un resultado», sin saber lo que ocurre realmente internamente, dejando preguntas sobre cómo se ponderan los datos, cómo se manipulan datos anómalos, qué variables se utilizan y si existen algunas más importantes que otras, entre otras. Este desconocimiento de cómo funcionan los algoritmos es lo que se considera opacidad, y se identifican tres grados de ella (Burrell, 2016):

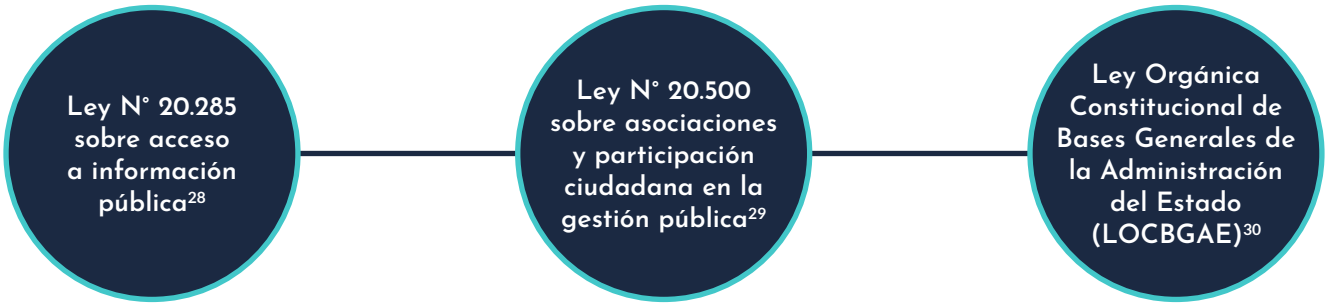
- **Opacidad intrínseca**: Surge cuando los algoritmos utilizan técnicas complejas de datos y no es posible explicar la relación entre ellos con el resultado, como es el caso de las redes neuronales.

- **Opacidad intencional**: Ocurre cuando no es deseable entregar información precisa de cómo funciona un algoritmo. Este nivel de opacidad puede ser deseable cuando se trata de algoritmos de fiscalización, en que el objetivo es lograr fiscalizaciones efectivas de manera proactiva, por lo cual no es deseable que los sujetos a fiscalización conozcan todos los detalles y logren burlar el sistema. También puede ocurrir cuando un algoritmo está protegido con propiedad intelectual.

- **Opacidad analfabeta**: Guarda relación con el poco conocimiento de la ciudadanía de cómo se entrenan y funcionan los algoritmos. Dado que los algoritmos son entrenados por equipos técnicos, y que el común de la población no necesariamente tiene el mismo nivel de conocimientos técnicos, se crea una brecha entre lo que explica el equipo técnico y lo que entiende la población general.

3.2.1
Normativa chilena

Leyes relevantes sobre transparencia y rendición de cuentas:



Más allá de las necesidades de transparencia algorítmica propias del proyecto, y su opacidad, el Estado de Chile posee una normativa en cuanto a transparencia: la Ley N° 20.285 sobre acceso a información pública. En particular, esta ley establece que «es pública la información elaborada con presupuesto público y toda otra información que obre en su poder», por lo cual sistemas de decisión que se financien con recursos públicos deben hacer transparente la información concerniente a ella de manera activa. Si bien la actual ley no establece de manera explícita la transparencia en algoritmos implementados en instituciones públicas, sí se debe considerar un mínimo de información cumpliendo con el principio de transparencia activa dispuesta en la ley.

Guardando relación con el punto anterior sobre el tratamiento de datos personales y sensibles, es importante destacar que el tratamiento de datos, como actividad estatal, debe ser transparente, lo que no implica publicar los datos objeto de tratamiento, pero sí las políticas, información, prácticas, protocolos y registros que se consideren, como la obligación de registro de la base de datos dispuesta en el artículo 22 de la Ley N° 19.628. Por tanto, dichos datos no se incorporan dentro de los antecedentes que deben ser puestos a disposición del público, resguardando así su privacidad por la Ley N° 19.628.

28 https://www.bcn.cl/leychile/navegar?idLey=20285
29 https://www.bcn.cl/leychile/navegar?idLey=20500
30 https://www.bcn.cl/leychile/navegar?idNorma=191865

La Ley N° 20.500, sobre asociaciones y participación ciudadana en la gestión pública, incorporó al Artículo 72 del DFL 1-19653 que fija el texto refundido, coordinado y sistematizado de la Ley Nº 18.575, Orgánica Constitucional de Bases Generales de la Administración del Estado, dice que «Los órganos de la Administración del Estado, anualmente, darán cuenta pública participativa a la ciudadanía de la gestión de sus políticas, planes, programas, acciones y de su ejecución presupuestaria (...)». En ese sentido, la creación de sistemas de decisión automáticos o semiautomáticos también deben estar sujetos a las cuentas públicas.

Esta misma norma dispone, dentro de los principios de la Administración del Estado, la transparencia y publicidad administrativa y participación ciudadana en la gestión pública, por lo cual los sistemas de decisión automatizadas también debieran estar sujetos a ello y debieran buscar que las instituciones comuniquen las implicancias de la nueva herramienta a la ciudadanía, que reciban retroalimentación y hagan las modificaciones necesarias para entregar mayor transparencia en la herramienta. Dentro de la misma ley, en el Título IV de la participación ciudadana en la gestión pública, el Aartículo 73 establece mecanismos de participación ciudadana para materias de interés ciudadano. Un algoritmo que se utilice para la entrega de beneficios sociales por supuesto que tiene un claro interés ciudadano. Más allá de lo dispuesto en la ley, entregar información y recibir comentarios de parte de la ciudadanía es un proceso beneficioso para la ejecución de algoritmos, considerando con especial interés a los actores que deberán interactuar con el modelo y que pueden entregar información relevante acerca de potenciales riesgos que pueden ocurrir en la etapa de implementación del mismo.

En la próxima sección se revisará el concepto de Licencia Social, que es la aceptación por parte de la ciudadanía de la implementación de una herramienta, aspecto clave para lograr su viabilidad.
Como se observa, la normativa antes citada configura obligaciones claras de transparencia y rendición de cuentas de los algoritmos que deben ser implementadas por las entidades públicas.

3.2.2
Consideraciones éticas: más allá de la ley

El nivel de opacidad determinará el nivel de transparencia de un algoritmo. Sin embargo, no porque un sistema sea opaco significa que no se debe entregar información pública sobre el mismo. Un concepto que se ha utilizado en ciencia de datos es el de «transparencia significativa», el cual refiere a la necesidad de comunicar los aspectos básicos de un algoritmo, poniendo especial cuidado en los impactos que tiene en la ciudadanía y lo que esta debe saber (Brauneis & Goodman, 2018). Brauneis y Goodman establecen los siguientes componentes de transparencia significativa:

- **Política del algoritmo**: Se refiere a cómo y bajo qué parámetros se toman las decisiones en la etapa de entrenamiento del algoritmo.

- **Rendimiento del algoritmo**: Se refiere a medidas de bondad de ajuste de un algoritmo, es decir qué tan bien es capaz de representar o predecir la realidad.

- **Equidad del algoritmo**: Se refiere a en qué medida el algoritmo se encuentra replicando sesgos. Una buena práctica es presentar resultados de la herramienta para distintos subgrupos de la población, por ejemplo hombres/mujeres, buscando que los resultados no difieran entre ellos.

- **Efectos del algoritmo en la capacidad del gobierno**: Se refiere a cómo se utiliza el algoritmo por parte de los funcionarios públicos.

Más allá de lo dispuesto por las leyes chilenas con respecto a información y transparencia, es deseable que existan mecanismos estables de comunicación hacia la ciudadanía y de manera interna sobre los aspectos claves del sistema de toma o soporte de decisión. Para realizar esta comunicación es importante que se realice con un lenguaje claro para disminuir la opacidad del instrumento, que sea fácil de encontrar o solicitar y que disponga de información de contacto. Es importante, por lo tanto, que se vayan documentando todas las decisiones desde la etapa de formulación, con tal de poder disponer de la mayor cantidad de información a la ciudadanía, que no ponga en riesgo el impacto del modelo. Una buena práctica es seguir los lineamientos de transparencia significativa.

Junto con esto también es importante determinar las responsabilidades y quiénes serán las personas y autoridades que serán las últimas responsables por el funcionamiento del modelo.

Preguntas claves:

- **¿Qué partes interesadas deberían estar al tanto del proyecto? Las partes interesadas suelen incluir a formuladores de políticas, trabajadores de primera línea, organizaciones de la sociedad civil, organismos públicos, personas que se verán afectadas por las acciones, etc.**
- **¿Has considerado algún mecanismo para que las partes interesadas se comuniquen con la institución por el proyecto?**
- **¿Será necesario explicar los mecanismos de toma de decisión o análisis a implementar? ¿Por qué? Evaluar el nivel de opacidad que corresponde a la herramienta. Si se transparenta todo el algoritmo, ¿se pone en riesgo su implementación?**
- **En caso de ocurrir un requerimiento de información respecto del proyecto, ¿quién es el encargado/a de elaborar la respuesta?**
- **¿Quién es responsable si el sistema se equivoca? Esto se refiere a tener clariad sobre la autoridad responsable y es lo que en inglés de denomina accountability.**
- **¿Tienes previstos mecanismos de monitoreo, control y evaluación? ¿Cómo se documentarán y qué periodicidad tendrán? ¿Se presentarán los resultados a la ciudadanía?**
- **¿Tienes previstos mecanismos de formación para comprender las responsabilidades, obligaciones legales y éticas entre el equipo participante?**

Buenas prácticas:

- Incluir a la ciudadanía en el debate de la necesidad, beneficios y riesgos de la implementación de dicha tecnología, entregando información oportuna y recibiendo comentarios que busquen mejorar y mitigar posibles riesgos de acuerdo a lo dispuesto en la Ley N° 20.500.

- Planificar mecanismos de comunicación para implementar luego en la etapa de implementación de la herramienta. Se aconseja que sea a través de los canales usuales de información de la institución con la ciudadanía, con lenguaje claro y de fácil acceso y navegación.

- Documentar todas las decisiones tomadas en el proceso de formulación e implementación de la herramienta que luego alimenten los mecanismos de comunicación anteriormente identificados. Para facilitar el proceso de documentación de decisiones, se puede utilizar la **Ficha de Diseño y Factibilidad del Proyecto** disponible en *Uso responsable de IA para política pública: Manual de formulación de proyectos* publicado por el Banco Interamericano de Desarrollo (Denis, Hermosilla, Aracena, Sánchez Ávalos, González Alarcón, & Pombo, 2021). Más adelante, una vez se comience la etapa de ejecución del proyecto se pueden utilizar las herramientas **Perfil del Modelo y Perfil de los datos** presentes en el mismo documento.[31]

- Identificar si corresponde crear un mecanismo de respuesta a solicitudes individuales y evaluar los canales más idóneos para realizarlo, junto con las personas que estarán a cargo de ellos.

31 Documento disponible en https://publications.iadb.org/es/uso-responsable-de-ia-para-politica-publica-manual-de-formulacion-de-proyectos

- Identificar la necesidad de transparentar lo más posible el algoritmo utilizado, procurando no afectar su implementación futura. Por ejemplo, en el caso de realizar un modelo de fiscalizaciones proactivas es poco aconsejable transparentar el código, e incluso las variables relevantes, ya que no sería deseable que las personas, en conocimiento del funcionamiento, alteren las decisiones de las fiscalizaciones. Por otro lado, cuando se trata de entrega de beneficios sociales, se puede explicar en mucho mayor detalle el algoritmo. Un caso concreto en Chile es el Sistema de Admisión Escolar, implementado por el Ministerio de Educación, que entrega información detallada de las variables claves del algoritmo de selección.

- En caso de ser posible, evaluar la utilización de softwares de código abierto (por ejemplo: Python, R).

3.3 Sesgos y discriminación

Los modelos de decisión automáticos pueden resultar sesgados, bien por el algoritmo mismo o porque la base de datos que se utiliza para su entrenamiento posee sesgos. Sesgo se puede entender como el error sistematizado que tiende a favorecer constantemente en una dirección. Es común pensar que los sesgos en un sistema automatizado se deben únicamente a la utilización de datos que ya se encuentran sesgados. Sin embargo, pueden existir múltiples razones que expliquen el sesgo en los sistemas (Hao, 2019):

- **Los datos se recopilan de manera sesgada**: Por ejemplo, si uno realiza una encuesta a la salida de las estaciones de Metro, dicha muestra puede ser significativa, pero únicamente de las personas que utilizan dicho transporte público, por lo cual no necesariamente será indicativa de la realidad nacional. Un ejemplo de esto en el ámbito público es cuando se realiza una encuesta para caracterizar hogares que pregunta por ocupación; una opción sesgada sería «dueña de casa» ya que «dueña» es femenino, por lo cual es poco probable que hombres, incluso

ejerciendo labores domésticas y de cuidado, se identifiquen dentro de esa categoría. Si los resultados de dicha encuesta son insumo importante para beneficios sociales, entonces este sesgo en la recopilación puede impactar directamente la entrega de estos.

- **Definición del problema**: Cuando se quiere implementar un sistema de decisión automático o semiautomático es porque se quiere resolver un problema específico, que puede ser de política pública o interno de la institución. En este primer paso de definición del problema pueden surgir sesgos si no se tiene una mirada global que incorpore a todas las personas que son impactadas por el problema, identificando específicamente si existen subgrupos de la población para los cuales se quiere garantizar equidad.

- **Se utilizan promedios sin tener en cuenta a subgrupos de la población**: Al entrenar un modelo, en la mayoría de los casos, se utilizan datos históricos y cada observación tiene el mismo peso. Esto podría traer un problema en cuanto puede favorecer a grupos de la población que son mayoría, dejando de lado las necesidades particulares de grupos minoritarios.

Una herramienta sesgada puede llevar a discriminar a usuarios y beneficiarios del sistema de decisión. Un ejemplo es lo que ocurrió con Amazon y su modelo automático de contrataciones, el cual favorecía constantemente a hombres postulantes, simplemente porque históricamente en la industria tecnológica los empleados eran hombres, por lo cual el modelo de IA discriminaba en contra de las mujeres. Al utilizar los datos de contrataciones de los últimos 10 años, la herramienta tendía a favorecer a hombres porque durante ese periodo la mayoría de las contrataciones fueron masculinas, y por lo tanto el sistema consideraba como factores indeseados si una persona había asistido a una universidad de solo mujeres o si una candidata había sido parte de un club de mujeres. El sistema no discriminaba directamente por género pero «aprendió» que características asociadas a postulantes mujeres no eran deseables en un candidato. Si bien la herramienta era un proyecto piloto y nunca se utilizó para tomar decisiones de contratación de manera automática, de todas formas fue finalmente retirada por la compañía en 2017 (Dastin, 2018).

Otro ejemplo de sesgos en los datos se encuentra en los sistemas de reconocimiento de imágenes. Un ejemplo es el caso de ImageNet, base de datos que entrega definiciones a imágenes a gran escala, información que luego es utilizada para entrenar modelos de reconocimiento facial, entendimiento de la actividad humana, entre otros. Uno de los principales problemas es la desigualdad de representación de imágenes dado un contexto. Por ejemplo, al buscar imágenes de reuniones de trabajo, los resultados son más bien imágenes de hombres blancos en torno a una mesa de reunión. Por lo cual, no se representa la diversidad que puede existir en un grupo de trabajo en cuanto a género, edad y racial. Por lo cual, este sesgo puede ser luego replicado al sistema a implementar que utiliza dicha categoría como base de datos (Yang, Qinami, Fei-Fei, Deng & Russakovsky, 2020).

En tecnologías de reconocimiento facial también se observan sesgos en los resultados del sistema que vienen dados por el entrenamiento del mismo. Para entrenar el modelo se utilizan mayoritariamente datos (imágenes) de hombres caucásicos, por lo cual los sistemas poseen tasas de errores mucho mayores en los demás grupos de la población (Buolamwini & Gebru, 2018).

Los sesgos en los datos pueden ser trabajados en el proceso de ejecución de un proyecto (antes de que sea desplegado para la población general), pero es relevante tenerlos en cuenta desde su etapa de formulación para buscar soluciones a potenciales efectos indeseados del sistema. Cuando se toman decisiones de política pública en base a sesgos necesariamente se discriminan a grupos de la población, lo cual está explícitamente prohibido en la ley chilena como se verá a continuación.

3.3.1
Normativa chilena

Leyes relevantes sobre sesgos y discriminación:

Ley N° 20.609 que establece medidas contra la discriminación[32]

32 https://www.bcn.cl/leychile/navegar?idLey=20609

La Constitución Chilena garantiza a todas las personas la igualdad ante la ley, estableciendo que en Chile no hay persona ni grupo privilegiado. Ni la ley ni autoridad alguna podrán establecer diferencias arbitrarias. También se garantiza la igual protección de la ley en el ejercicio de los derechos de las personas y un debido proceso, esto es, un procedimiento justo.

Es importante destacar que Chile también ha suscrito tratados internacionales en cuanto a la no discriminación, como por ejemplo la Convención Internacional sobre eliminación de todas las formas de discriminación racial a través del decreto 747[33] y el Pacto Internacional de Derechos Civiles y Políticos[34].

La legislación chilena actual, por medio de la Ley N° 20.609, establece la no discriminación arbitraria y consagra un «mecanismo judicial que permite restablecer eficazmente el imperio del derecho toda vez que se cometa un acto de discriminación arbitraria». Dicha ley define discriminación arbitraria como «toda distinción, exclusión o restricción que carezca de justificación razonable». Por lo cual, la presencia de sesgos en un sistema de soporte o toma de decisión no es deseable no solamente por el concepto de justicia algorítmica sino también por las implicancias legales que puede tener.

La ley indica las categorías protegidas que deberán entonces ser consideradas en las evaluaciones sobre sesgo algorítmico. Estas categorías son: la raza o etnia, la nacionalidad, la situación socioeconómica, el idioma, la ideología u opinión política, la religión o creencia, la sindicación o participación en organizaciones gremiales o la falta de ellas, el sexo, la maternidad, la lactancia materna, el amamantamiento, la orientación sexual, la identidad y expresión de género, el estado civil, la edad, la filiación, la apariencia personal y la enfermedad o discapacidad.

Si el modelo entrenado resulta que arroja resultados sesgados que perjudican a personas individuales, entonces se deben tener en cuenta las medidas que establece la Ley para dichos casos. Una buena práctica es contar con un canal exclusivo que dé razones sobre los resultados del algoritmo y que permita ingresar solicitudes de rectificación.

Es importante recordar que, si bien los sistemas automáticos ahorran tiempo y recursos en tomar decisiones, éstos son entrenados por personas, por lo cual, al igual que las personas, pueden cometer errores. Lo importante es contar con un protocolo definido para esos casos de ocurrencia.

33 https://www.bcn.cl/leychile/navegar?idNorma=400589

34 https://bibliotecadigital.indh.cl/handle/123456789/1020#:~:text=Adoptado%20por%20la%20Asamblea%20General,los%20pueblos%2C%20el%20goce%20de

3.3.2
Consideraciones éticas: más allá de la ley

Si bien la Ley N° 20.609 antidiscriminación prohíbe la discriminación arbitraria en Chile, es importante que, a la hora de formular un proyecto, el equipo responsable y quien lidere, comprenda aspectos mínimos que puede impactar la justicia del algoritmo a desarrollar. Como se comentó anteriormente, los sesgos en un modelo pueden tener distintas fuentes.

Incluso si se realiza un proceso de compra pública para tercerizar la creación de la herramienta, las personas a cargo del proyecto en la institución deben tener en cuenta los sesgos que se pueden generar, para comunicarlos adecuadamente al equipo técnico a cargo del modelamiento, estableciendo los estándares de equidad necesarios y deseables.

Es relevante destacar que una de las fuentes de sesgos guarda relación con la forma en que se describe el problema. Las personas dentro de una institución tienen mayor conocimiento del problema que se quiere solucionar y es, por lo tanto, clave que se comunique claramente a quienes estarán a cargo del desarrollo del modelo aspectos claves que pueden resultar en discriminaciones: ¿Existen subgrupos de la población a los cuales se les quiera asegurar equidad? ¿Existen desigualdades en el proceso a intervenir? (Denis, Hermosilla, Aracena, Sánchez Ávalos, González Alarcón & Pombo, 2021).

En caso de que el modelo entregue resultados sesgados será importante que se tomen en cuenta medidas de respuesta y mitigación para casos personales, que las personas perjudicadas por el algoritmo puedan ingresar reclamos y se entregue una respuesta individualizada a su solicitud. Es importante, por lo tanto, determinar un equipo responsable de este tipo de comunicaciones.

Preguntas claves:

- **¿Qué inequidades de base hay en el proceso/entorno donde se inserta el proyecto? Por ejemplo, si se requiere de acceso a internet**
- **¿Existen grupos específicos (vulnerables) para los que deseas garantizar la equidad de los resultados o la protección de sus derechos? Por ejemplo, grupos dado su género, edad, localización, clase social, nivel educativo, urbano-rural, etnia**
- **¿Qué sesgos crees que podrían tener los datos?**

Buenas prácticas:

- Identificar en el mapeo inicial de datos si existen sesgos en los datos que se utilizarán en el modelamiento. En caso de ser así, identificar posibles medidas de mitigación que podrían estar en el modelo mismo o bien a través de mecanismos de rectificación ex-post para resultados indeseados. Tener en cuenta estos potenciales sesgos y evaluarlos en detalle en la etapa de ejecución con el equipo técnico, solicitando información estadística detallada para subgrupos de la población (media, desviación estándar, distribución, entre otras).
- Entender la naturaleza de datos atípicos presentes en la data. Entregar contexto sobre dichas observaciones para luego decidir cómo se tratarán, buscando minimizar efectos negativos en el resultado del modelo o herramienta.

- Planificar la necesidad de realizar auditorías de sesgos al algoritmo de forma periódica. Por ejemplo, la herramienta Aequitas, auditoría de sesgos de código abierto, permite evaluar distintas métricas de paridad estadística como la paridad en falsos positivos y en falsos negativos (http://www.datasciencepublicpolicy.org/our-work/tools-guides/aequitas/). El uso de herramientas de este tipo sirve para analizar si el modelo entrenado es equitativo. También es una buena práctica realizar auditorías algorítmicas, las cuales pueden hacerse de manera interna o externa.

- Crear un mecanismo de respuesta a solicitudes individuales.

- Discutir internamente acerca del problema que se quiere solucionar, incluyendo a diversos actores relevantes, para evitar sesgos en su definición.

3.4 Otras consideraciones éticas

Más allá de las consideraciones legales que deben tener en cuenta quienes formulen proyectos de ciencia de datos, descritas en el inciso anterior, existen consideraciones éticas que también deben estar presentes en este proceso.

- **Proporcionalidad:** Es importante al formular un proyecto de ciencia de datos, identificar si existen otras soluciones no basadas en tecnología que pueden dar respuesta al problema identificado. El desarrollo de modelos de ciencia de datos puede ser costoso en recursos y tiempo, por lo cual es necesario realizar un análisis costo beneficio de la posible solución. Además, como ya se presentó en este documento, puede tener efectos adversos en la ciudadanía por lo que es de especial importancia comunicar el por qué la solución tomada es la mejor para solucionar el problema, identificando claramente los beneficios

y riesgos potenciales. En el estudio del Derecho existen tres subprincipios de proporcionalidad que también pueden ser aplicados a los sistemas objeto de esta guía:

- **Análisis de idoneidad:** Se refiere a que la solución efectivamente tenga el efecto deseado en resolver el problema.

- **Análisis de necesidad:** Se refiere a que la solución es la mejor para la persona. En el caso del desarrollo de sistemas de toma o soporte de decisión, esto nace de responder la pregunta ¿existe otra forma en la que se puede dar solución al problema?

- **Proporcionalidad estricta:** Se refiere a que los beneficios de la solución implementada sean mayores que los potenciales riesgos.

El principio de proporcionalidad, y su debida comunicación a la ciudadanía es clave para lograr la aceptación general de la herramienta (licencia social).

Licencia social: La licencia social se entiende como la aceptación de la implementación de la herramienta por parte de la ciudadanía (Data Futures Partnership, 2017). Alcanzar licencia social no es trivial y es clave para que el proyecto sea viable ya que, en caso de que no se logre licencia social, se corre el riesgo de que el proyecto finalmente nunca se implemente. Muchas veces no bastará con simplemente cumplir con las disposiciones legales, sino que se deberá cumplir mayores estándares de comunicación.

Existen distintas formas de alcanzar la licencia social, pero lo principal tiene que ver con la comunicación clara de los beneficios y riesgos de la herramienta a la ciudadanía (Eticas, 2021). Esto se vincula directamente con lo expuesto anteriormente sobre transparencia y rendición de cuentas. Una buena práctica es que las organizaciones posean una plataforma de información con un lenguaje claro que entregue los conceptos más relevantes del algoritmo.

Data Futures Partnership, grupo independiente neozelandés designado por el gobierno que busca impulsar el uso confiable de datos y fortalecer el ecosistema de datos de Nueva Zelanda, establece ocho preguntas claves en tres categorías que deben ser respondidas por organizaciones que utilizarán datos personales para alcanzar licencia social (Data Futures Partnership, 2021).

- **Qué:** ¿Para qué será utilizada mi información?
- **Quién:** ¿Cuáles son los beneficios y quiénes serán los beneficiarios? ¿Quién va a utilizar mis datos?
- **Cómo:** ¿Están mis datos seguros? ¿Se anonimizarán mis datos? ¿Puedo ver y corregir mis datos? ¿Se me pedirá Consentimiento Informado? ¿Podrían vender mis datos?

Además de las leyes listadas en esta sección, se deberán tomar en cuenta también leyes sectoriales según sea la naturaleza del proyecto e institución. Por ejemplo, un proyecto en el área de salud debe tener en cuenta la Ley N° 20.584 de derechos y deberes del paciente[35]; un proyecto sobre reclamos de consumidores tendrá que tener en cuenta la Ley N° 19.496 del consumidor[36]. Por lo tanto, será tarea del equipo a cargo del sistema evaluar si éste se enmarca dentro de sus atribuciones legales.

35 https://www.bcn.cl/leychile/navegar?idLey=20584
36 https://www.bcn.cl/leychile/navegar?idLey=19496

Preguntas claves:

- **¿Crees que un sistema de ciencia de datos/IA es el medio adecuado para resolver el problema? ¿Por qué? ¿Has evaluado otras alternativas?**
- **¿Qué impactos negativos podría tener tu proyecto? Revisa casos de uso similares.**
- **¿Crees que los usuarios/afectados encontrarán aceptable el uso de datos planteado para resolver el problema? ¿Por qué?**
- **Si la población completa del país se entera de tu proyecto, ¿lo aprobará? ¿Por qué?**

Buenas prácticas:

- Comparar casos de uso similares en la región y el mundo, para así dimensionar la proporcionalidad, los impactos y la aceptación de la solución propuesta teniendo en cuenta los costos y beneficios y potenciales riesgos que podría tener el desarrollo del modelo.
- Favorecer la transparencia activa del algoritmo en canales de comunicación activa.
- Promover procesos de participación acerca del sistema de decisión para evaluar su licencia social.
- Involucrar a todas las áreas en que el sistema pudiera impactar, por ejemplo, el área legal de la institución sobre otras posibles normativas que podrían impactar la ejecución de la herramienta, el área usuaria, las áreas encargadas de participación ciudadana, entre otros.

04

Casos de uso y riesgos legales y éticos

4.1 Sistemas de reconocimiento y de detección de eventos

4.2 Predicción

4.3 Personalización

4.4 Soporte de interacción

4.5 Optimización

4.6 Razonamiento con estructuras de conocimiento

4. Casos de uso y riesgos legales y éticos

Para ilustrar las principales consideraciones legales y éticas, se presentan a continuación algunos casos de uso de la IA nacionales, según clasificación de tareas de la OECD (OECD, 2022). Cada uno de los riesgos legales y éticos aquí descritos pueden ser mitigados utilizando las buenas prácticas estipuladas en el Capítulo 3 de esta guía para cada consideración.

Estos casos de uso se obtuvieron del repositorio de algoritmos públicos desarrollado por el GobLab UAI. Es importante destacar que las herramientas de ciencia de datos pueden cumplir distintas tareas a la vez, pero en esta guía se ha simplificado identificando aquella tarea principal.

Los ejemplos aquí presentados son una muestra de los presentes en la página web. Para conocer más acerca del repositorio y de los criterios que deben cumplir los proyectos subidos en la página puede ingresar a https://algoritmospublicos.cl/.

4.1

Sistemas de reconocimiento y de detección de eventos

Se refiere a la capacidad de reconocer o detectar una situación que está ocurriendo en tiempo presente. Por un lado, los sistemas de reconocimiento se refieren a la identificación y categorización de datos en forma de video, imagen, texto, etc. El objetivo de este tipo de análisis es poder identificar a una persona, situación u objeto mediante el análisis de fotos, videos, u otros. Uno de los usos más conocidos de esta tecnología en el día de hoy son los desbloqueos de aparatos electrónicos con reconocimiento facial, o la identificación automática de personas en redes sociales, aunque se ha ido restringiendo esta última aplicación.[37]

Por otro lado, la detección de eventos busca visualizar un estado de realidad actual a partir de bases de datos estructuradas, poniendo énfasis en la detección de patrones y anomalías en tiempo presente.

37 https://about.fb.com/news/2021/11/update-on-use-of-face-recognition/
38 https://www.consejotransparencia.cl/cplt-entrega-recomendaciones-para-que-municipios-utilicen-correctamente-drones-y-globos-de-televigilancia/

Principales consideraciones éticas

Uso de información personal y sensible:

Tratamiento y seguridad. En casos de identificación de personas es claro el uso de datos personales sensibles, en cuanto los datos biométricos son considerados como tales. Por lo cual, habrá que regirse por la Ley de protección de la vida privada. Más allá de las consideraciones legales, se debe tener especial cuidado en el almacenamiento y protección que se le entreguen a los mismos datos.

Por otro lado, en casos de televigilancia y de utilización de drones en que no se tratan datos biométricos de todas maneras se debe tener cuidado con las imágenes captadas y su uso. El Consejo para la Transparencia (CPLT) entregó en 2017 una serie de recomendaciones para guiar a municipalidades que opten por este tipo de tecnología.[38]

Sesgos y discriminación:

En Estados Unidos se ha documentado que los errores en sistemas de reconocimiento facial son más grandes en mujeres que en hombres, y en minorías raciales, específicamente en personas afrodescendientes (Lohr, 2018; Garvie & Frankle, 2016). Incluso si la data con la que se entrena el modelo no presenta sesgos, es importante realizar un análisis de disparidades sobre el desempeño, para identificar resultados sesgados. Si bien en algunos casos de reconocimiento facial, como el desbloqueo de un aparato electrónico, puede no impactar mucho el sesgo del sistema, esta conclusión cambia si hablamos de detenciones o decisiones de registro en aeropuertos, en cuanto el algoritmo está replicando sesgos históricamente perjudiciales para minorías raciales. Esto podría tener incluso efectos en los derechos de libertad de movimiento, la libertad de manifestación pacífica o la libertad de expresión, entre otros.

Transparencia y rendición de cuentas:

En particular, todos los modelos de reconocimiento y detección deben siempre contar con sistemas de respuesta a solicitudes particulares. Como se argumentó en la sección de datos personales y en la sección de sesgos, es importante que exista un sistema donde las personas puedan ingresar sus reclamos al uso de la herramienta y recibir una rectificación o restitución acorde.

En Chile podemos identificar los siguientes casos de tecnologías de reconocimiento y detección de eventos:

1 **Reconocimiento facial:**

En la municipalidad de Las Condes se instalaron cámaras de reconocimiento facial que toman una fotografía del rostro de una persona para luego analizar los datos biométricos, buscando con ello identificar a dicha persona. Luego, esa identidad se cruza con la base de datos de la PDI para ver si esa persona posee causas pendientes y proceder con la detención. Nótese que en este caso los datos biométricos de una persona, que son datos sensibles, se utilizan para poder individualizar a un sujeto, por lo cual la consideración de datos personales y sensibles guarda relación especial con la forma en la que se trata y se almacena la información. Es de suma importancia en casos como estos contar con un alto nivel de transparencia, haciendo hincapié en los potenciales beneficios para la ciudadanía, para así lograr licencia social.

2 DART-Teledx:

Detección de retinopatía diabética mediante tamizaje automatizado de imágenes de fondo de ojo. Este proyecto comenzó como un proyecto piloto en 2016 y actualmente ya se utiliza en más de 140 establecimientos públicos de salud, realizando más de 350.000 análisis de exámenes. Los datos utilizados, imágenes de fondo de ojo, son datos sensibles, en cuanto entregan información respecto de la salud de una persona. Al igual que en el ejemplo anterior, es importante el resguardo de la privacidad de las personas. En este caso, además, se debe tomar en consideración la Ley N° 20.584 de derechos y deberes del paciente.

3 Fiscalización de vías exclusivas y pistas sólo bus con cámaras automatizadas:

Actualmente se fiscaliza el tránsito por vías exclusivas por medio de reconocimiento de fotografías de placas patentes, las cuales se sitúan a una distancia de dos o tres cuadras, buscando así los casos en que conductores utilizan la vía exclusiva más allá de lo permitido. A partir de este reconocimiento y con la validación por parte de un agente público del Programa Nacional de Fiscalización se cursa la multa correspondiente.

4 Analista virtual en gestión de licencias médicas:

FONASA implementó este sistema que permite clasificar automáticamente las licencias en aquellas que deben ser aceptadas y aquellas que requieren revisión de un médico contralor. Claramente utiliza datos sensibles, en cuanto el estado de salud de una persona lo es. Es importante transparentar, aunque manteniendo un cierto grado de opacidad, algunos criterios que resultan en el rechazo de la licencia médica y atender solicitudes personales para rectificar el resultado del algoritmo en caso de ser necesario.

4.2
Predicción

Se refiere a la predicción de un estado futuro en base a información pasadas. Se buscará predecir el valor de una variable en base a variables conocidas.

Principales consideraciones éticas:

Sesgos y discriminación:

Una de las fuentes de sesgos puede estar en que los mismos datos lo estén. Si se utilizan datos sesgados para realizar predicciones, los resultados también lo estarán.

Transparencia y rendición de cuentas:

Dependiendo de la tecnología utilizada en su modelamiento, los modelos predictivos pueden ser muy difíciles de entender. Es importante que la institución entregue información al menos sobre los beneficios del sistema y cómo se utiliza actualmente. En los casos de predicción de infracciones, es importante mantener un cierto grado de opacidad para evitar que aquellos infractores encuentren nuevas formas de evadir la fiscalización.

Licencia social:

En modelos de predicción, la licencia social es un punto clave sobre todo si está asociado a la entrega de beneficios o una acción institucional en particular. Será importante comunicar claramente cómo se toman las decisiones y si el modelo es un sistema de toma o de soporte de decisión. Puede ser recomendable incluir el factor humano para acercar la herramienta a la ciudadanía. Es relevante también contar con sistemas de información directa hacia aquellas personas o instituciones que se podrían ver afectadas por la implementación del modelo.

En Chile podemos identificar los siguientes casos de uso de esta tecnología:

1 Predicción de riesgo de deserción escolar:

El Ministerio de Educación y el Ministerio de Desarrollo Social y Familia desarrollaron un sistema que utiliza datos administrativos para predecir el riesgo de deserción escolar de niños y niñas de entre 7mo básico y 4to medio. Los establecimientos educacionales acceden a la información sobre este riesgo de deserción para así focalizar sus esfuerzos en aquellos estudiantes con más riesgo. En este caso, se utilizan datos personales de los niños y niñas y de la conformación de su hogar, por lo que hay que tener especial cuidado en no generar discriminación por estos mismos factores. En este caso en particular, se trata de un sistema de soporte a la toma de decisión que informa el accionar que tomarán los establecimientos educacionales, lo cual suele entregar mayor confianza en la aplicación de la herramienta.

2 Modelo predictivo para fiscalizaciones proactivas:

La Dirección del Trabajo utiliza un modelo predictivo de infracciones desarrollado por la Universidad de Chile. Este modelo consiste en fiscalizaciones preventivas a empresas con mayor riesgo de infracción sin necesidad de esperar una denuncia previa. Un ejemplo de sesgos a partir de datos históricos se puede dar en este caso si existió una empresa con muchas denuncias pasadas pero que no ha tenido infracciones en las fiscalizaciones proactivas dado, por ejemplo, a un cambio en los protocolos. Dado este posible efecto negativo, es importante que las empresas estén al tanto de la implementación del modelo y establecer canales de comunicación. Sin embargo, en este caso es relevante mantener un cierto grado de opacidad, ya que no es deseable que se conozca el detalle completo del funcionamiento del modelo.

3 Sistema Alerta Niñez:

El ministerio de Desarrollo Social está realizando el piloto de este proyecto, el cual utiliza información de niños, niñas y adolescentes (NNA) para generar un sistema de alertas, por presentar situaciones que podrían dar paso a una vulneración de derechos. Las Oficinas Locales de Niñez (OLN) contactan a la familia y, si aceptan la invitación, se trabaja en conjunto en un plan que incluye el apoyo y acompañamiento para fortalecer herramientas que posee la familia para la crianza de los NNA, así como también para el acceso a los servicios y programas que ofrece el Estado en la comuna. En caso de ser necesario, también se ofrece un servicio de terapia familiar para apoyar la resolución de problemáticas en la dinámica de la familia. En este caso se están utilizando datos personales y datos personales sensibles, por lo que se deben tener los resguardos de información requeridas para el tratamiento de dicho tipo de datos. Además, es importante no generar discriminación en las intervenciones realizadas por las OLN.

4.3
Personalización

Se refiere al desarrollo de perfiles de usuarios que, con base en los datos generados por sus propias acciones, vayan mejorando en el tiempo.

Principales consideraciones éticas:

Consentimiento informado:

En los casos en que la implementación de un modelo de personalización esté fuera de las competencias legales de la institución será necesario contar con el consentimiento informado de los usuarios. Independiente de la disposición legal, siempre es bueno contar con consentimiento en cuanto constituye una base de licitud para la aplicación de la herramienta.

Discriminación:

Este tipo de herramientas puede entrar en conflicto con la libertad de las personas y puede producir sesgos, en cuanto tiende a recomendar contenido que las personas ya han consumido previamente y no toma en cuenta nuevos intereses. Este concepto es algo a lo cual la ciudadanía está bastante acostumbrada en las recomendaciones de películas o series de plataformas de streaming. Sin embargo, la aplicación una tecnología de este tipo por parte del gobierno puede resultar más problemática debido al principio de no discriminación. Una forma de reducir este riesgo sería que las personas tengan acceso fácil al catálogo completo de acciones.

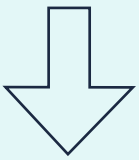
En el repositorio de algoritmos públicos no existen hasta el momento proyectos de esta naturaleza, pero se podría pensar, por ejemplo, en proyectos que recomienden cursos de perfeccionamiento. Es importante en este caso contar con el consentimiento de las personas que participen del o no circunscribir los cursos disponibles solamente a aquellos que se asemejan o son cercanos a los intereses previos de la persona. El Servicio Nacional del Consumidor, SERNAC, se refiere a este desafío como "Resguardo de la libertad de elección", que es parte de la ley de protección del consumidor.[39]

39 https://www.sernac.cl/portal/618/articles-64740_archivo_01.pdf

4.4
Soporte de Interacción

Se refiere al apoyo en la interacción entre humanos y máquinas, pueden ser de manera escrita o por voz. Se busca que la herramienta dé respuestas ad-hoc a lo vertido por el usuario, a través de protocolos predeterminados por experiencia pasada.

Principales consideraciones éticas:



Transparencia:

En casos de asistentes virtuales, es de suma importancia que las personas sepan que están tratando con una máquina y no con una persona. Dado que el sistema se basa en protocolos, no siempre dará una respuesta satisfactoria, lo cual puede llevar a un sentimiento de frustración por parte del usuario.

Uso y protección de información:

Las consideraciones de uso de datos personales y sensibles tienen dos aristas en este caso.

a. Información utilizada para entrenar el modelo: Para entrenar un asistente virtual se utilizan interacciones pasadas, historial en que puede existir tanto información personal como sensible. El tratamiento de dichos datos debe hacerse conforme a la ley.

b. Información vertida en el asistente virtual: Se deben tener consideraciones adicionales de seguridad con respecto a la información personal y sensible, en cuanto debe estar protegida de ataques cibernéticos.

En Chile podemos identificar los siguientes casos de uso de esta tecnología:

1 WhatsApp Mujer:

El Ministerio de la Mujer e Igualdad de Género y el Servicio Nacional de la Mujer y Equidad de Género, desarrollaron en 2020 un asistente virtual que permite administrar múltiples conversaciones y determina si la persona debe ser derivada a un asistente especializado. La necesidad de esta herramienta surge a raíz de las cuarentenas por Covid-19, periodo en que una parte importante de mujeres agredidas o en riesgo no podían comunicarse de manera telefónica por estar compartiendo el espacio con su agresor. Esta herramienta, dado su objetivo, debe tener especial cuidado en el uso y protección de los datos personales y sensibles vertidos en el asistente.

2 Asistente virtual FOSIS:

Responde preguntas que tienen los postulantes a FOSIS. El asistente fue entrenado con preguntas frecuentes y aprende sobre variaciones en el lenguaje. En caso de no comprender una pregunta, la conversación es derivada con un ejecutivo. Esta política de derivación automática va en línea con la transparencia y busca evitar la frustración de que el asistente no entienda lo que uno está preguntando.

4.5
Optimización

Se refiere a un proceso de optimización de proceso realizado de manera automática. El objetivo es que el algoritmo aprenda a través de la simulación de escenarios.

Principales consideraciones éticas:

Transparencia algorítmica y rendición de cuentas:

Especialmente si el proceso de optimización tiene un impacto directo en la ciudadanía, es importante que se comuniquen los factores que se toman en cuenta en el algoritmo de decisión. Conjuntamente, es importante comunicar los beneficios que tiene la implementación de la herramienta, poniendo en relieve el bien común por sobre la experiencia personal. Adicionalmente, es importante poder contar con un canal de atención directo.

Uso de información personal y sensible:

El algoritmo será entrenado con datos, por lo cual se puede estar en presencia de datos personales y sensibles. Habrá que regirse entonces por la ley y dar la protección necesaria, según corresponda. También es relevante evaluar si los datos fueron entregados con la finalidad del proyecto que se quiere realizar.

En Chile podemos identificar los siguientes casos de uso de esta tecnología:

1 Algoritmo de asignación del Sistema de Admisión Escolar:

En 2016 comenzó a implementarse este algoritmo para priorizar la inscripción de alumnos en establecimientos educacionales según las preferencias de apoderados y la información sobre el núcleo familiar. Si bien este sistema es un algoritmo de decisión que está basado en reglas establecidas por ley, no ha estado exento de polémicas, ya que cada cierto tiempo aparecen experiencias negativas con la herramienta. Es de suma importancia entonces la existencia de canales de reclamo directos para evaluar casos particulares y rectificar los resultados del algoritmo en caso de un error.

2 Algoritmo para priorizar listas de espera no GES:

El Ministerio de Salud comenzó en 2018 la implementación del sistema de priorización de listas de espera no GES en base al riesgo de la persona. Para esto debió primero calcularse el riesgo, tomando como datos criterios clínicos explícitos, como son la presencia de tumores malignos, condiciones crónicas, uso de medicamentos, edad y sexo, entre otros. Todos esos datos utilizados son datos personales y sensibles que el Ministerio puede utilizar por estar dentro de su competencia legal. Esta es una herramienta de soporte de decisión y son los médicos quienes toman la decisión final de la entrega de hora de consulta.

4.6

Razonamiento con estructuras de conocimiento

Se refiere a la posibilidad de establecer relaciones causales mediante simulaciones de eventos que no han ocurrido aún. La diferencia con el análisis predictivo es, primero, la relación causal y, en segundo lugar, la no existencia previa del estado futuro. Suele utilizarse en áreas de la salud y legal.

Principales consideraciones éticas:

Sesgos y discriminación:

Al igual que en los modelos predictivos, el uso de datos sesgados puede llevar a resultados sesgados. Un ejemplo de sesgos en este tipo de tecnología en Estados Unidos es una herramienta de predicción de reincidencia de delitos en personas que ya han sido procesadas por delitos. Históricamente este país registra más detenciones de hombres que de mujeres y mayormente de minorías raciales, especialmente afrodescendientes. Por lo tanto, este modelo entrenado con esos datos sesgados evaluaba peor a hombres de minorías raciales, lo cual se traducía en una condena más dura (Angwin, Larson, Mattu & Kirchner, 2016).

Uso de información personal y sensible:

El algoritmo será entrenado con datos, por lo cual se puede estar en presencia de datos personales y sensibles. Habrá que regirse entonces por la ley y dar la protección necesaria según corresponda.

En Chile podemos identificar el siguiente caso de uso de esta tecnología:

1. **Aplicación que usa inteligencia artificial para audiencias de control de detención:** La Defensoría Penal Pública (DPP) se encuentra pilotando una herramienta que crea perfiles de imputados y apoya la estrategia de la defensa proyectando posibles salidas o medidas cautelares en base a casos pasados similares. Dado que el fallo de medidas cautelares no ha pasado anteriormente, esta herramienta es de razonamiento. Claramente en este caso se utilizan datos personales no solamente del imputado/a sino también de casos pasados y es importante por lo tanto cumplir con todas las medidas de tratamiento y seguridad descritas en la ley.

Esta lista de ejemplos es meramente ilustradora y no es exhaustiva de todos los proyectos actualmente en implementación en Chile. Así también, pueden existir otros riesgos éticos y legales que no se encuentran en esta guía pero que pueden afectar al proyecto.

05

Conclusiones

5. Conclusiones

Al momento de formular un proyecto de IA o de ciencia de datos, no debes olvidar:

Referencias

AEPD, Agencia Española de Protección de Datos (2021). Medidas de cumplimiento. disponible en : https://www.aepd.es/es/derechos-y-deberes/cumple-tus-deberes/medidas-de-cumplimiento

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks.* ProPublica. Disponible en: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Brauneis, R. and Goodman, E.P. (2018) *Algorithmic Transparency for the Smart City.* Yale Journal of Law & Technology, 20, 103. https://doi.org/10.31228/osf.io/fjhw8

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on fairness, accountability and transparency (pp. 77-91). PMLR.

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society, 3(1), 2053951715622512.

Dastin, J. (2018). *Amazon Scraps Secret AI Recruiting Tool that Showed Bias Against Women.* Disponible en https://www.reuters.com/article/us-amazon-com-jobs-automationinsight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-womenidUSKCN1MK08G Consultado el 03 de noviembre de 2021.

Data Futures Partnership (2017). *A Path to Social Licence: Guidelines for Trusted Data Use.*

Data Futures Partnership (2021). *Trusted Data Use Guidelines For Aotearoa New Zealand.*

Denis, G., Hermosilla, M., Aracena, C., Sánchez Ávalos, R., González Alarcón, N., & Pombo, C. (2021). *Uso responsable de IA para política pública: Manual de Formulación de Proyectos.*

Eticas Consulting (2021). *Guía de Auditoría Algorítmica.* Disponible en: https://www.eticasconsulting.com/eticas-consulting-guia-de-auditoria-algoritmica-para-desarrollar-algoritmos-justos-y-eficaces/

Garvie, C., & Frankle, J. (2016). *Facial-recognition software might have a racial bias problem.* The Atlantic, 7.

Hao, K. (2019). *This is how AI bias really happens—and why it's so hard to fix. MIT Technology Review.*

Ischen, C., Araujo, T., Voorveld, H., van Noort, G., & Smit, E. (2019, November). *Privacy concerns in chatbot interactions.* In International Workshop on Chatbot Research and Design (pp. 34-48). Springer, Cham.

OECD (2022), "OECD Framework for the Classification of AI systems", OECD Digital Economy Papers, No. 323, OECD Publishing, Paris, https://doi.org/10.1787/cb6d9eca-en

Lohr, S. (2018). *Facial recognition is accurate, if you're a white guy.* New York Times, 9(8), 283.

González, F., Ortiz, T., & Sánchez, R. (2020). *IA Responsable.*

Red Iberoamericana de Protección de datos personales (2017). *Estándares de protección de datos personales.* Disponible en https://www.redipd.org/es/documentos/estandares-iberoamericanos

Yang, K., Qinami, K., Fei-Fei, L., Deng, J., & Russakovsky, O. (2020). Towards fairer datasets: Filtering and balancing the distribution of the people subtree in the imagenet hierarchy. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 547-558).

Anexo 1: Glosario

Algoritmo: Conjunto de instrucciones sobre cómo una computadora debe realizar una tarea en particular. Son utilizados por muchas organizaciones para tomar decisiones y asignar recursos basados en grandes conjuntos de datos. Se comparan con las recetas, que toman un conjunto específico de ingredientes y los transforman a través de una serie de pasos explicables en un resultado predecible (Caplan, Donovan, Hanson, & Matthews, 2018).

Base de datos: Una pieza de software que almacena datos de un tipo particular en un formato que admite el acceso de baja latencia. Es un marco para almacenar y acceder de manera eficiente a los datos. Una base de datos prototípica es un servidor fornido que contiene más datos de los que cabría en una computadora normal, los almacena de manera que se pueda acceder rápidamente (esto generalmente implica una tonelada de optimizaciones ocultas para que el usuario la pueda usar más fácilmente), y está preparado para recibir solicitudes de otras computadoras para acceder o modificar los datos (Cady, 2017).

Ciencia de datos: Se define comúnmente como una metodología mediante la cual se pueden inferir ideas procesables a partir de los datos. La realización de la ciencia de datos es una tarea con un objetivo ambicioso: la producción de creencias informadas por datos y para ser utilizadas como base para la toma de decisiones (Igual & Seguí, 2017).

Consentimiento: Toda manifestación de voluntad libre, específica, informada e inequívoca por la que el titular acepta, ya sea mediante una declaración o una clara acción afirmativa, el tratamiento de datos personales que le conciernen (Boletín N° 11.092-07, 2017).

Datos estructurados: Datos altamente organizados y fáciles de descifrar por algoritmos. Por ejemplo: fechas, nombres, direcciones, entre otras (IBM, 2021).

Datos no estructurados: Datos que no pueden ser procesados ni analizados con herramientas de análisis tradicionales. Por ejemplo: texto, actividad redes sociales, entre otras (IBM, 2021).

Entrenamiento de un modelo: Es la fase de la ciencia de datos donde se intenta ajustar de la mejor manera el algoritmo en base a los datos disponibles (C3.ai).

Licencia social: Describe cómo las expectativas de la sociedad con respecto a algunas actividades pueden ir más allá del cumplimiento de los requisitos de la regulación formal; aquellos que no cumplan las condiciones para la licencia social (incluso si cumplen formalmente) pueden experimentar desafíos y cuestionamientos continuos. Son las expectativas de la sociedad con respecto a la conducta y actividades de las corporaciones, que van más allá de los requisitos de la regulación formal (Carter, Laurie & Dixon-Woods, 2015).

Sesgo (estadística): Posible consecuencia de negar a determinados miembros de la población la oportunidad de ser seleccionados para la muestra. Como resultado, la muestra puede no ser representativa de la población (Lind, Marchall & Wathen, 2005).

Sesgo (sistema computacional): Sistemas informáticos que discriminan de forma sistemática e injusta a determinadas personas o grupos de individuos a favor de otros. Un sistema discrimina injustamente si niega una oportunidad o un bien o si asigna un resultado indeseable a un individuo o grupo de individuos por motivos que no son razonables o apropiados (Friedman & Nissenbaum, 1996).

Referencias Glosario

Boletín N° 11.092-07 y 11.144-07 (2017). Proyecto de Ley refundidos sobre protección de datos personales.

C3.ai. Glossary. Model Training. disponible en: https://c3.ai/glossary/data-science/model-training/

Cady, F. (2017). The data science handbook. John WiLey & Sons.

Caplan, R., Donovan, J., Hanson, L., & Matthews, J. (2018). Algorithmic accountability: A primer. Data & Society, 18.

Carter, P., Laurie, G. T., & Dixon-Woods, M. (2015). The social licence for research: why care. data ran into trouble. Journal of medical ethics, 41(5), 404-409.

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. ACM Transactions on Information Systems (TOIS), 14(3), 330-347.

IBM Cloud Education (2021). Structured vs. Unstructured Data: What's the Difference?

Igual, L., & Seguí, S. (2017). Introduction to Data Science. In Introduction to Data Science (pp. 1-4). Springer, Cham.

Stock, J. H., Watson, M. W., & Larrión, R. S. (2012). Introducción a la Econometría (No. 330.1543 S8). Pearson.

GUÍA

Formulación ética de proyectos de ciencia de datos

BID LAB
UNIVERSIDAD ADOLFO IBÁÑEZ
gob digital CHILE