

# CIENCIA DE DATOS y ABOGADOS

# INTRODUCCION

## Abogados y DS's:

- Desde la perspectiva de la regulación legal y ética aplicable al desempeño de DS.
- Desde la interacción entre abogados y DS's en actividades de estos últimos.
- Desde la visión del uso de DS como herramienta en las tareas propias de un abogado en el ejercicio de su profesión.

# ASPECTOS A ABORDAR



Diferentes visiones y perspectivas sobre DS.



Aspectos éticos y legales que involucra el ejercicio de DS, principalmente en lo relativo al manejo de big data y la IA.



Casos prácticos



¿Cómo pueden trabajar mejor juntos?

# Diferentes perspectivas

## DS

- **Optimización y Rendimiento del Modelo:** Busca modelos precisos y eficientes. Su prioridad es precisión y rendimiento, aunque una mayor complejidad dificulte la explicabilidad.

## ABOGADO

- **Énfasis en Protección de Derechos Individuales:** Respetar privacidad y no discriminación. Garantizar que el manejo de datos cumpla leyes de protección de datos y respeten derechos de las personas.

# Diferentes perspectivas

## DS

- **Exploración de Datos y Descubrimiento de Patrones:** Descubrir patrones y tendencias ocultas en los datos para obtener información valiosa. Análisis exhaustivo de datos, incluso si hay riesgos potenciales de sesgo o discriminación.

## ABOGADO

- **Cumplimiento Normativo y Responsabilidad Legal:** Cumplir normativas legales, como RGPD y LOPD. Minimizar riesgo de demandas y litigios relacionados con privacidad y uso indebido de datos.

# Diferentes perspectivas

## DS

- **Balance entre Precisión y Equidad:** Desafío es equilibrar precisión del modelo con equidad y la ausencia de sesgo. Resistencia a sacrificar precisión en nombre de equidad, más si consideran ciertos sesgos como inherentes a los datos.

## ABOGADO

**Énfasis en Transparencia y Explicabilidad:** Sobre todo en procesos de toma de decisiones automatizadas, para evitar demandas por decisiones injustas o discriminatorias. Asegurar que los usuarios comprendan cómo se están utilizando sus datos.


# Diferentes perspectivas

## DS

- **Innovación y Avance Tecnológico:** Atención principal en la innovación y el avance de tecnología. Menos interés en aspectos legales y éticos, más en resultados.

## ABOGADO

**Foco en Responsabilidad y Diligencia Debida:** Contratos claros entre las partes involucradas en proyectos de DS. Definir roles y responsabilidades para mitigar problemas legales.



## Implicaciones legales y éticas en DS:

*El problema no es la tecnología  
sino el uso que se haga de ella.*





# Tratamiento de datos

---



# Tratamiento de datos. Marco legal

- Reglamento General para la Protección de Datos (RGPD de 2016).
- Ley Orgánica de Protección de Datos y Garantía de los Derechos Digitales (LOPD de 2018). Remite en su mayor parte al RGPD.
- Ambas normativas establece requisitos y responsabilidades en materia de protección de datos en empresas y la mayor parte del sector público, así como los derechos que asisten a usuarios y consumidores.

# Consideraciones RGPD

- Dirigido fundamentalmente a tomar conciencia de todas las empresas que almacenan nuestros datos y permitirnos decidir informadamente sobre esto.
- Políticas de privacidad: Impone rotundamente que deben ser claras y entendibles para personas normales, no expertos legales. Exige implementar estrategias de información y transparencia por capas, garantizando un mínimo indispensable de información visible.

# Consideraciones LOPD.../1

- Adapta el RGPD al ámbito nacional, no puede contravenirlo.
- Garantiza y protege el tratamiento de datos personales y derechos fundamentales de las personas físicas, especialmente derecho al honor, a la privacidad personal y a la seguridad.
- El ciudadano debe ser capaz de controlar su propia información y que pueda tomar decisiones sobre ésta, como quién puede utilizarla, para qué, hasta cuando, si puede compartirse, etc. También decidir si a ella se puede acceder total o parcialmente.

# Consideraciones LOPD.../2

- Agencia Española de Protección de Datos. No basta inscribir el "fichero". Se sustituye por un "Registro de Actividades de Tratamientos" (RAT), que además debe describir las medidas de seguridad aplicables en cada caso.
- Aplica a toda persona u organización que trate total o parcialmente automatizado de datos personales. Se exceptúan casos como "actividades exclusivamente personales o domésticas". Ejm no aplica a la agenda de contactos personales del tlf., pero sí a una sección de comentarios de un blog, que recoge nombre y dirección o email.

# Consideraciones LOPD.../3

- Aplica a todas las entidades públicas o privadas, pequeños profesionales, pymes, etc.
- Define al "responsable" como la persona que decide con qué finalidad se usan esos datos, con quién se comparten, qué tiempo, etc. Ejm. si tienes una web con un formulario de suscripción, tú eres el responsable.
- Datos especialmente protegidos: ideología, religión, creencias, origen racial, salud, vida sexual, comisión o infracciones penales o administrativas. También añadidos datos genéticos, biométricos y sobre orientación sexual.



# Consideraciones LOPD.../4

- Exige condiciones como la seudonimización o anonimización, el cifrado, informes de impacto sobre la privacidad.
- Obligación de recabar el consentimiento explícito y específico para una finalidad concreta.
- Se han incrementado en un 33% hasta 2022, las denuncias sobre estos temas.
- Sanciones muy graves, graves y leves. Multa Max 20M € o 4% facturación bruta anual. Aparte indemnización si se demuestra el daño.

# Derechos en la LOPD.../1

- **Derecho de acceso:** Saber si un responsable tiene datos nuestros, cuáles son, qué han hecho con ellos y si los han enviado a alguien.
- **Derecho de rectificación:** Cuando en un "fichero" hay datos nuestros no exactos o completos (se viola principio de exactitud).
- **Derecho de supresión:** Posibilidad de pedir eliminación de un fichero en ciertos supuestos. Ejm. Ya no son necesarios, usos no autorizados, retiro de consentimiento.



# Derechos de la LOPD.../2

- **Derecho a la limitación del tratamiento:** Posibilidad de paralizar el uso de datos por causas específicas. Ejm. mientras un tribunal decide una impugnación de un fichero por datos incorrectos.
- **Derecho a la portabilidad:** Se puede pedir a un responsable que transfiera nuestros datos a otro responsable. Ejm. de una operadora telefónica a otra.
- **Derecho de oposición:** a) por mal tratamiento de nuestros datos y b) por perjuicios a nuestros derechos derivados de decisiones automatizadas. Ejm. En banca, el scoring de clientes en función de ciertos patrones para decidir sobre un préstamo, cuando éstas se basan únicamente en esos perfiles

# Ética en el tratamiento de datos.../1

- **Ética:** Conjunto de principios consensuados en una comunidad que dirigen y valoran como bueno o malo el comportamiento. Ejm. No dañar, respetar, ser justo.
- La ética en este contexto implica consideraciones tanto para los individuos cuyos datos se manejan como para la sociedad en general. Su sentido es garantizar que los datos se traten de manera justa, responsable y respetuosa, evitando daños innecesarios y promoviendo el bienestar de las personas y la integridad de la información

# Ética en el tratamiento de datos.../2

- **Privacidad:** Respetar el derecho de las personas a tener control sobre su información personal y evitar su divulgación no autorizada.
- **Consentimiento Informado:** Obtener el permiso explícito y comprensible de los individuos antes de recopilar o utilizar sus datos.
- **Transparencia:** Ser claro y transparente sobre cómo se recopilan, utilizan y procesan los datos, y brindar información clara sobre los propósitos y las implicaciones.
- **Equidad:** Evitar la discriminación y asegurarse de que los datos y los algoritmos sean justos y no perpetúen sesgos.
- **Minimización de Datos:** Recolectar solo los datos necesarios y relevantes para el propósito establecido y evitar la recopilación excesiva.
- **Exactitud:** Garantizar la precisión y la calidad de los datos recopilados y procesados.

# Ética en el tratamiento de datos.../3

- **seguridad:** Proteger los datos contra accesos no autorizados, pérdida, robo o manipulación.
- **Integridad:** Asegurarse de que los datos no sean manipulados o alterados de manera indebida.
- **Responsabilidad:** Ser responsable de las decisiones y acciones relacionadas con los datos y estar preparado para enfrentar las consecuencias de manera ética.
- **Cumplimiento Legal:** Respetar las leyes y regulaciones aplicables al tratamiento de datos.
- **Beneficio Social:** Evaluar cómo el tratamiento de datos puede contribuir positivamente a la sociedad y minimizar los posibles daños.
- **Educación y Sensibilización:** Promover la educación sobre la importancia de la ética en el tratamiento de datos entre los profesionales y el público en general.

# Riesgo de divulgación. ../1

Se refiere a la posibilidad de que la información sensible o confidencial se revele involuntariamente a través de los resultados, análisis o visualizaciones compartidas en el proceso de divulgación.

Esto puede ocurrir cuando se comparten datos o resultados de manera que los individuos o entidades involucrados en el proyecto puedan ser identificados, a pesar de que los datos hayan sido anonimizados o desidentificados previamente.

El riesgo de divulgación en ciencia de datos es particularmente relevante en proyectos que involucren datos personales o confidenciales, como información médica, financiera o de identificación personal.

# Riesgo de divulgación .../2

- Según el RGPD, datos personales significa cualquier información relacionada con una persona física identificada o identificable("sujeto de datos"). Persona física identificable es aquella que puede identificarse, directa o indirectamente, por referencia a un identificador como un nombre, número de identificación, datos de ubicación, identificador en línea o a uno o más factores específicos del estado físico, fisiológico, identidad genética, psíquica, económica, cultural o social de esa persona física (atributos).
- Los controladores de datos y los que los procesan, tienen la responsabilidad legal de evitar la divulgación de la identidad de los interesados, pudiendo dar lugar a multas y litigios. Aquí cobran relevancia el consentimiento y la privacidad.

# Riesgo de divulgación .../3

## Tipos:

- Divulgación de identidad: Puede ser directa o indirecta. Ejm. en una encuesta de un pueblo, muestra una sola dentista negra, se sabrá quién es ella.
- Revelación de atributos: Se revela información de esa persona que antes no se sabía. El daño depende del atributo expuesto.
- Revelación inferencial: Ejm. Por datos de tablas se desprenden o infieren otras circunstancias del sujeto de datos
- Divulgación de población o de un equipo: Ejm. Se establece relación entre las características de un determinado rango de empleados y el salario de un individuo.

# Riesgo de divulgación.../4

Salidas "seguras" de datos y "datos abiertos":

- Se distinguen salidas seguras de datos, donde es más fácil manejar que no ocurra la reidentificación, porque están más controlados. Ejm. Con datos tradicionales. Estos solo están disponibles después de una anonimización cuidadosa para que no haya divulgación.
- Para datos abiertos ("no seguros") fuera de una instalación de datos segura, en general se piensa que no hay riesgo de reidentificación porque son datos anónimos, pero aquí el enfoque es que la reidentificación es de rompecabezas, ya que la multiplicidad de información permite armar la imagen completa de un individuo.





"La IA será más relevante y supondrá un cambio más profundo que el fuego, la electricidad o internet"

CEO Google Sundar Pichai

# Inteligencia Artificial

# Inteligencia Artificial. Ideas sobre el marco legal

- No existe una regulación sistemática nacional o internacional en IA.
- La UE lidera intentos de consenso de los principios y normas éticas esenciales.
- Uso de regulación contractual y autorregulación.
- En España se incluyó un artículo (23) en la Ley para la Igualdad de Trato y la No Discriminación, dedicado a la IA y a los mecanismos para la toma de decisiones automatizadas. Allí se prevén obligaciones para las Adm. Pcas. y empresas de tomar medidas de minimización de sesgos en los algoritmos y cumplir con la transparencia y rendición de cuentas "siempre que sea factible técnicamente".

# Inteligencia Artificial. Directrices

- RGPD e iniciativas específicas de la Comisión Europea. Actualmente hay dos propuestas(2020 y 2021), que han sufrido múltiples modificaciones.
- Libro blanco de IA de la Unión Europea (2020): vela por los derechos de los consumidores y la promoción de la innovación en el ámbito de la IA. Expresa la visión y el enfoque de la UE en el desarrollo y regulación de la IA en un marco para fomentar su adopción de manera ética, responsable y competitiva. Son estrategias y directrices generales que requieren regulaciones posteriores.

# Ética en Inteligencia Artificial

Principios éticos básicos:

- Control y supervisión humana.
- Confianza, seguridad, robustez, fiabilidad.
- Privacidad y gobernanza de datos.
- Transparencia y explicabilidad.
- Respeto de derechos humanos (Esp. Libertad, dignidad y autonomía).
- No discriminación (igualdad, equidad, justicia).
- Sostenibilidad ambiental.
- Responsabilidad proactiva y rendición de cuentas.

Otros como precaución, accesibilidad y reversibilidad.



# Sesgo .../1

Sucede cuando un sistema de aprendizaje automático refleja los valores de las personas que lo desarrollaron o entrenaron. Hablamos de prejuicios, concepciones de la realidad por las cuales tomamos decisiones de forma inconsciente. Son predisposiciones psicológicas que nos hacen llegar a unas determinadas conclusiones de manera automática.

## Sesgo.../2

Caso sobre *la contratación experimental de Amazon* hace unos años. Empleaba una herramienta con IA para encontrar candidatos clasificándolos de una a cinco estrellas, parecida a como califican sus productos. Era discriminatoria para las mujeres porque los modelos informáticos de Amazon estaban entrenados para analizar a los candidatos mediante la observación de patrones en los CV's enviados a la empresa durante un período de 10 años y al estar basado en los empleados actuales que eran mayoritariamente hombres, el algoritmo fijó su criterio de candidato ideal como masculino, penalizando cualquier solicitud que contuviera la palabra mujer.



## Consideraciones sobre IA

La ausencia de un sistema legal eficiente para la regulación de la IA hace que la autorregulación asuma un rol protagónico. Las políticas y orientaciones éticas y de cada organización o responsable son guías para el uso correcto y confiable de la IA.

Los abogados, en conjunto con los DS, tienen la tarea de crear los marcos contractuales de las empresas, entidades o responsables que regirán ese uso en los términos allí señalados, eso sí con validez sólo entre las partes.

Es primordial que en esos contratos se prevean y cuiden todos los aspectos y circunstancias posibles de manera minuciosa.

# Casos prácticos.

1. Creación de equipos de DS's y abogados en trabajo ordinario de un científico de datos.
2. Cómo influye el marco ético y legal en los profesionales de data en el sector público en Países Bajos.





# 1. DS Y LEY. Aprendiendo juntos IA responsable.../1

Varios autores de Universidades Internacionales. Febrero, 2022

- El Consejo legal de una red social pidió al equipo DS, asegurar que su sistema lograra equilibrio entre la necesidad de eliminar contenido incitador y la libertad de expresión. Se detectó que la definición de "contenido incitador" era muy simplista y lamentaron no haber trabajado desde un principio conjuntamente con el equipo legal. A pesar de usar aparentemente el mismo lenguaje, mucho se perdía en la traducción y programación.
- Los sistemas creados deben ser "responsables" e incorporar valores como justicia, equidad, responsabilidad y privacidad. Ello requiere una estrecha colaboración entre DS's y abogados.

# DS Y LEY. Aprendiendo juntos IA responsable .../2

Se establecieron objetivos como:

- Diálogo multidisciplinario: identificando conceptos, términos, perspectivas, etc . para lograr un lenguaje compartido
- Marco de actuación de la IA responsable: considerar el impacto en las personas, como temas éticos, efectos no deseados y daños potenciales en todas las fases del ciclo.
- Responsabilidad profesional: superar la idea de que ésta recae exclusivamente en los abogados, sino lograr una visión compartida de lo que es una IA responsable.

## DS Y LEY. Aprendiendo juntos IA responsable.... /3

Se diseñó un curso especial para ayudar a promover la colaboración entre equipos de DS y abogados, como experimento para ver la dinámica práctica en tareas conjuntas para resolver problemas en el mundo real de los retos de IA.

Se centró en temas como responsabilidad legal, responsabilidad general, discriminación e igualdad, transparencia y privacidad.

Se plantearon 6 sesiones de 4 horas al mes, con 20 estudiantes de cada equipo, en la Universidad de Tel Aviv.

# DS Y LEY. Aprendiendo juntos IA responsable.... /4

## Conclusiones:

- En la fase más técnica hubo incomodidad entre ambos equipos. Algunos quisieron abandonar la experiencia por el choque de visiones de cada lado, sobre todo los de DS, en el simulacro sobre la explicabilidad de un sistema de calificación de créditos, donde asumían roles de regulador y regulado. Los abogados indicaron que nunca se habían sentido tan frustrados, por primera vez no entendían nada.
- En la fase menos intensa sobre posicionamiento disciplinario, se les pidió hacer un balance entre dchos. humanos y su compensación con IA, y los límites en el diseño de sistemas. Se detectó que los DS's se enfocaban más en lo concreto y preciso en las definiciones, mientras que los abogados eran más especulativos, con más tendencia a las interpretaciones o posibles múltiples significados.

## 2. Cómo los marcos éticos y legales influyen en los profesionales de data en el sector público en Países Bajos.../1

Muchos escándalos sobre el mal uso y prácticas abusivas en Holanda han propiciado debates entre el público y la comunidad de DS, sobre la manera correcta de hacer DS.

Cada vez más el sector público trabaja con procesos de data centralizada y sistemas de algoritmos insertos en el área del gobierno público.

Se encargó un estudio para analizar esta realidad en el ámbito de oficiales públicos, para quienes hay altos estándares de exigencia. Se enfoca en dos entidades del sector público de los Países Bajos:

1. Municipalidades
2. Policía Nacional

Se parte de la aseveración de que los practicantes de DS y otros profesionales de la data deberían ser responsables de las decisiones éticas. Ello supone conocer realmente con qué tipo de decisiones deben lidiar estos en su trabajo ordinario.

# Cómo los marcos éticos y legales influyen en los profesionales de data en el sector público en Países Bajos..../2

Es un estudio interpretativo.

Países Bajos posee un sistema robusto de centralización y automatización de datos que hace presumir que cumple con todos los marcos regulatorios aplicables.

1. Los municipios (son 352 pero se toman 25 para este estudio), cuentan con cierta independencia por la delegación de muchas tareas, como registrar ciudadanos para temas de beneficencia a desempleados, albergues, etc.
2. La policía nacional es una fuerza descentralizada que tiene algunos sistemas o redes a nivel de proyectos entre sus dependencias locales y regionales. Desarrolla sus propias tecnologías de algoritmos y tienen más conocimiento en este campo que en los municipios.

Se empleó la observación directa de los empleados en sus tareas cotidianas y otros planos como entrevistas, interacciones, etc, en un rango de 4 años (2017-2021).

# Cómo los marcos éticos y legales influyen en los profesionales de data en el sector público en Países Bajos.... /3

## Marco de estudio analítico:

- El foco del estudio está orientado por la propuesta de que "la responsabilidad de servir a la sociedad a través de un buen manejo y trabajo de la data recae en los practicantes o profesionales de la data, en todos ellos".
- Estos practicantes no tienen per se un conocimiento especial sobre normas o marco legal al respecto, sino que para sus decisiones emplearán su criterio, ética personal y normas profesionales.
- El límite entre marco legal y ético es difuso o borroso entre algoritmos e IA.

# Cómo los marcos éticos y legales influyen en los profesionales de data en el sector público en Países Bajos.... /4

- El estudio concluye que las declaraciones sobre ética y regulaciones legales de IA no se cumplen en su mayoría y no son seguidas en la práctica de estos trabajadores. El límite entre ambos marcos es difuso o borroso
- Muchas veces el conocimiento profundo de los marcos está limitado a un número selecto de profesionales y este no se comparte con el resto del equipo
- La mayoría de los profesionales de data de los casos estudiados dicen que no usan marcos específicos de forma regular en su trabajo diario, sino que se valen de su relativo conocimiento particular. Sienten que ese cumplimiento no es su responsabilidad en primera instancia, sino que es el usuario o contratante de dichos sistemas el que debe asegurarse de cumplir con las condiciones legales previas antes de solicitar sus servicios.



# Cómo los marcos éticos y legales influyen en los profesionales de data en el sector público en Países Bajos.... /5

## Otras percepciones:

- Los profesionales de data generalmente son conscientes de que existen normas y reglas que rigen DS y la mayoría lo ve positivo. Incluso consideran que hay muchas "áreas grises" y convendría más legislación.
- Comprenden la distinción entre normas legales y las que no lo son (Ejm. éticas, técnicas, empíricas), pero en algunos casos la ambigüedad les plantea dudas.
- Muchos profesionales de data se sienten frustrados por las limitaciones a su trabajo, cuando hay normas restrictivas, pero tienden a aceptarlas. Ejm la decisión de la CE de no utilizar datos biométricos. La existencia de cámaras por doquier facilitaría mucho su trabajo, pero atenta contra la privacidad.

# Cómo los marcos éticos y legales influyen en los profesionales de data en el sector público en Países Bajos.... /6

Esta desconexión considerable entre los marcos y la práctica de los data pros, no es por desinterés o resistencia de estos, sino por la dinámica de falta de tiempo y capacidad para cumplir con esos extremos .

Se finaliza con la consideración sobre la necesidad de mayor alfabetización en esta área y una comprensión más profunda de las cuestiones e implicaciones legales y éticas de la DS en el sector público.

Actualmente los data pros tienen mucha discrecionalidad, debido en parte a que no hay muchos actores que tengan un nivel de alfabetización similar en estos temas.

Desarrollar la interfaz entre la ley y las prácticas profesionales sigue siendo clave para garantizar la implementación efectiva de los marcos éticos.

# CÓMO PUEDEN TRABAJAR MEJOR LOS ABOGADOS CON LOS CIENTIFICOS DE DATOS

# Sugerencias para una mejor comunicación.../1

- Dejar claro qué fuentes de datos están fuera de los límites o son altamente sensibles. Ejm. las variables de etnicidad, género y edad son inapropiadas para modelos de créditos bancarios, pero pueden ser apropiadas para modelos de asuntos médicos y de salud .
- Determinar de manera conjunta cómo la selección de data y el modelo de "tuneo" ayuda a cumplir ambos objetivos.
- Pueden crearse sistemas que desarrollen modelos que integren ambas visiones y objetivos funcionales, o sea el balance entre "precisión" y "equidad".

## Sugerencias para una mejor comunicación.../2

- Colaborar para construir y emplear herramientas que permitan automatizar la auditoría del abogado sobre el modelo. "No puedes manejar lo que no mides". Algunas de esas revisiones pueden ser cuantificadas, la equidad puede medirse.
- Puede haber sinergia. Las herramientas creadas para monitorear y potenciar la equidad pueden ayudar a mejorar la precisión. Ejm. "Explainable AI (xAI)" es un set de herramientas que se usa en DS para "iluminar la caja negra" de los modelos de machine learning. Se emplea para investigar qué variables son importantes para medir la precisión de sus modelos, pero también sirve para conocer qué variables conducen decisiones injustas.

## Sugerencias para una mejor comunicación.../3

- Negocios, gobiernos, investigadores y la sociedad en general, necesitan modelos que sean éticos y a la vez poderosos para proveer conocimientos valiosos al mismo tiempo. La única manera de lograrlo es trabajando estrechamente profesionales de data y abogados, comunicando la importancia de los objetivos-funciones de cada disciplina y buscando formas de optimizar ambos cometidos.