

Capítulo 6

Protección jurídica de las bases de datos para el entrenamiento de Redes Neuronales Artificiales (RNN) e inteligencia artificial (IA) en México con relación al Tratado de Libre Comercio entre México, Estados Unidos y Canadá (T-MEC)

José Zócimo Orozco Orozco¹

DOI: <https://doi.org/10.61728/AE26001906>



¹ Es Profesor docente e investigador en el Centro Universitario de Ciencias Sociales y Humanidades de la Universidad de Guadalajara desde hace más de cuatro décadas, donde imparte materias en la Licenciatura en Derecho, Maestría y Doctorado; miembro del Sistema Nacional de Investigadores Nivel I; tiene Perfil Prodep; es miembro de la Sociedad de Geografía y Estadística del Estado de Jalisco. Autor de los libros, entre ellos, *Historia del Derecho Universal y Mexicano*, Porrúa, 6^a ed., 2024; *Historia de la cultura jurídica tapatía*, 2016; coordinador de obras en las áreas de la Historia del derecho y Derecho intelectual, entre otras. Miembro del Cuerpo Académico Desafíos de la Propiedad Intelectual UDG 1115.

Resumen

En la primera parte se abarcan conceptos básicos acerca de las bases de datos para el entrenamiento de redes neuronales artificiales e inteligencia artificial (IA). La segunda parte aborda la discusión de la problemática sobre la protección de estas bases de datos, las cuales pueden contener datos no originales. Entre las principales cuestiones está si las bases de datos a que se refiere el presente artículo cuentan con protección jurídica y quién es el titular de los derechos. El presente artículo compara la legislación europea, líder en el ramo de la protección intelectual y la creación del derecho sui generis para bases de datos. Por último, se hace un estudio que pone de manifiesto si la legislación mexicana cuenta con alguna figura de protección similar al derecho sui generis y si está contemplada además en el Tratado de Libre Comercio entre México, Estados Unidos y Canadá (T-MEC).

I. Introducción

Las bases de datos son fundamentales para el entrenamiento de redes neuronales artificiales o RNN (por sus siglas en inglés). Este estudio aborda conceptos básicos acerca de las bases de datos para el entrenamiento de redes neuronales artificiales y su legislación en México con respecto al Tratado de Libre Comercio con Estados Unidos y Canadá (T-MEC).

Trata el problema de si la regulación jurídica aplicada a las bases de datos que se encuentra en el Tratado de Libre Comercio entre México, Estados Unidos y Canadá (T-MEC), acerca de las mencionadas bases de datos, es suficiente para tratar la problemática en cuestión. El presente estudio abarca el derecho de propiedad intelectual de las bases de datos para entrenar redes neuronales artificiales e IA y el Tratado de Libre Comercio entre México, Estados Unidos y Canadá (T-MEC).

La regulación de las bases de datos que entrena los grandes modelos de lenguaje y otros tipos de IA viene principalmente de la zona europea,

que ya cuenta con legislación en el tema, por lo que es de suma importancia observar la forma de abordar la problemática en México para hacer un esfuerzo por legislar o mejorar la legislación respectiva en la problemática actual.

II. Las Redes Neuronales Artificiales ANN

Basogain señala que las Redes Neuronales Artificiales ANN (Artificial Neural Networks), que están inspiradas en las redes neuronales biológicas del cerebro humano, están constituidas en elementos que se comportan de forma similar a la neurona biológica en sus funciones más comunes. Las ANN adquieren el conocimiento de una cosa por medio del estudio, ejercicio o experiencia.²

La neurona artificial

La pieza básica sobre la que se desarrollan las redes neuronales es la neurona artificial (o simplemente neurona), la cual puede ser entendida como una función que intenta modelar el funcionamiento de una neurona cerebral humana. La característica de flexibilidad de las redes neuronales está en su capacidad para concatenarse con otras neuronas artificiales.³

Isasi Viñuela, Pedro y Galván León destacan que la parte más importante de una red de neuronas artificiales es el aprendizaje. El esquema de aprendizaje de una red es lo que determina el tipo de problemas que será capaz de resolver. La capacidad para resolver un problema estará en función de los ejemplos de que dispone en el proceso de aprendizaje.

El aprendizaje debe ser significativo (haber un número suficiente de ejemplos) y ser representativo (los elementos del conjunto de aprendizaje deben de abarcar un conjunto amplio de ejemplos de aplicación).⁴

² Basogain, Xavier, *Redes Neuronales Artificiales*, (Escuela Superior de Ingeniería de Bilbao), 2008), p.2.

³ Sierra Ramos, Juan Miguel, López Montes, Antonio, *Introducción a las Redes Neuronales Artificiales*, Universidad Complutense de Madrid, 2022, p.8.

⁴ Isasi Viñuela, Pedro y Galván León, Inés, *Redes Neuronales Artificiales, un Enfoque Práctico*, México, Pearson, 2004, p.10.

El aprendizaje no supervisado

Moreno Pallares reafirma que las redes neuronales artificiales (ANN) con aprendizaje no supervisado, conocidas como auto supervisadas, pueden ajustar los pesos de las conexiones entre sus neuronas, sin necesidad de influencia externa. Para detener el aprendizaje es necesario que se cuente con presupuestos de detención como lo son:

- Cuando el cálculo del error sobre los ejemplos de entrenamiento ha alcanzado un mínimo;
- Cuando para cada uno de los ejemplos el error observado está por debajo de un límite.

Otra condición de detención del aprendizaje puede incluir un número predeterminado de ciclos de entrenamiento.⁵

Bases de datos para el aprendizaje de la IA

Como señala Raraz Vidal, las bases de datos son un pilar para el entrenamiento y desarrollo de modelos de IA. Estos depósitos de información alimentan la IA, les permiten aprender, adaptarse y tomar decisiones inteligentes. Las calidades de las bases de datos son fundamentales cuando se determina el éxito o fracaso de la IA; estos datos deben ser de calidad y representativos; además, deben ser precisos, imparciales y relevantes para la tarea que se vaya a realizar. A medida que las IA se vuelven más avanzadas, crece exponencialmente la demanda de datos.⁶

⁵ Moreno Pallares, Mario, Moreno Pallares, Rigoberto. Entrenamiento de Redes Neuronales Artificiales con Aprendizaje no Supervisado en el Control de Calidad del Análisis. *Revista Polo del Conocimiento* (Edición núm. Vol.7, No. 5. Mayo 2022, p.1587 DOI 10.23857/pc.v7i5.4048).

⁶ Raraz Vidal, Jarvis, La importancia de las Bases de datos para el Entrenamiento en Inteligencia Artificial. *Revista de Investigación en Salud* (2023):121. DOI 10.35839/repis.7.3.1970.

Los LLM Large Language Models (Grandes Modelos de Lenguaje)

Quezada Peralta (2025) hace referencia a que los grandes modelos de lenguaje son sistemas de inteligencia artificial especializados que procesan y generan texto de manera coherente. Entienden el lenguaje natural reconociendo grandes volúmenes de texto y predicen la palabra o secuencias de palabras siguientes. Se entrena con conjuntos de datos masivos que pueden incluir audio, video y cualquier texto disponible.⁷

Bases de datos

Fernández Iglesias define las bases de datos de la siguiente forma: son compilaciones de información organizada, almacenada y gestionada mediante un sistema informático. Las bases de datos son diseñadas para gestionar grandes volúmenes de datos y proporcionan una forma de acceder y almacenar los datos para procesarlos.⁸

Ventajas de las bases de datos

- Compactación. Al estar los archivos en digital, ahorran espacio.
- Velocidad. Las consultas pueden hacerse con rapidez sin necesidad de beneficios del enfoque de bases de datos.

C.J. Date en su libro *Introducción a las Bases de Datos* expresa:

Los datos pueden compartirse. No solo puede compartirse el uso de las bases de datos sino también desarrollar aplicaciones para operar sobre los mismos datos.

- Es posible reducir la redundancia;
- Es posible unir archivos para evitar que los archivos sean más grandes;

⁷ Quezada Peralta, Gonzalo, Donate López Juan, Abreu González, Rodrigo, *Análisis Comparativo de los Grandes Modelos de Lenguaje*, Monografías Secoir, consultado el 2 de julio de 2025, 501, <https://secoir.org/inteligencia-artificial-generativa/> .

⁸ Fernández Iglesias, José Manuel. *Pequeña Introducción a las Bases de Datos*. Universidad de Vigo, 2023, 1, DOI: 10.13140/RG.2.2.13924.65925

- Es posible (hasta cierto grado) evitar la inconsistencia. Si un dato está duplicado y uno se actualiza y otro no habrá inconsistencia;
- Es posible brindar un manejo de transacciones. Una transacción es una unidad de trabajo lógica como por ejemplo actualizar un dato A y un dato B;
- Es posible mantener la integridad. La integridad se refiere a que los datos de la base estén correctos;
- Es posible hacer cumplir la seguridad. Las restricciones de seguridad serán verificadas siempre que se intente acceder a datos sensibles;
- Es posible equilibrar los requerimientos en conflicto; y
- Es posible almacenar los datos y tener acceso más rápido a las aplicaciones más importantes a cambio de un acceso más lento a otras aplicaciones⁹.

Bases de datos y clasificación de datos

Schiaffino Silvia menciona: En la operación de estar clasificando los datos, el objetivo es darles una categoría a los datos en clases diferentes.

Primero se crea un modelo que se basa en la distribución de datos; luego el modelo se usa para clasificar nuevos datos. Una vez hecho el modelo, se puede predecir la clase de un nuevo dato.¹⁰

La predicción en los datos

El objetivo de la predicción es deducir el valor de futuros datos basándose en los atributos de otros valores.

Crear un modelo IA basándose en la forma como se distribuyen los datos.

Con el modelo podemos predecir valores que desconocemos o valores futuros.

⁹ Date C.J., *Introducción a los Sistemas de Bases de Datos* (Prentice Hall, 2001), 16-19

¹⁰ Schiaffino Silvia. *Inteligencia Artificial, Técnicas de Clasificación*. Universidad Nacional del Centro de Provincia de Buenos Aires, (2018), Recuperado de: https://users.exa.unicen.edu.ar/catedras/optia/public_html/clasificacion.pdf

Tipos de bases de datos

Gómez Fuentes, sobre las bases de datos hace distinción:

Bases de datos jerárquicas. Se apoyan sobre un conjunto de estructuras de árbol cuyos nodos hacen una representación de entidades de información y las uniones representan las relaciones;

Base de datos de red. Se representan sobre una estructura y un conjunto.

Bases de datos relacionales. Este modelo representa a los datos y las relaciones entre los datos con ayuda de una colección de tablas, y un numero de columnas con nombres únicos; y

Bases de datos orientadas a objetos. Estas bases de datos permiten especificar la estructura de objetos complejos. Además, proveen una identidad que es única para cada objeto el cual es independiente. Los objetos complejos pueden constituirse a partir de objetos más simples.¹¹

Los conceptos básicos del diseño orientado a objetos son:

- La abstracción. Se toman detalles relevantes de un objeto y se pasan al mundo virtual.
- El encapsulamiento. Es el proceso por el cual se ocultan detalles de un objeto que son innecesarios o esenciales para su uso.
- La herencia. Ayuda con la organización de un sistema clasificando los datos abstraídos por niveles.
- El polimorfismo. Es la propiedad que tiene un objeto de un tipo de mostrarse como de un tipo distinto.

Otros tipos de bases

- Bases de datos distribuidas. Son bases de datos que pueden estar conectadas a internet, con lo cual el usuario puede acceder a los datos de otra computadora como si estuviera en su propia computadora.
- Bases de datos objeto-relacionales. Comparten las características de las bases de datos relacionales y las bases de datos orientadas a objetos.

¹¹ Gómez Fuentes María del Carmen, Notas del Curso Bases de Datos, (Universidad Autónoma Metropolitana, México, 2013), 16-19.

- Bases de datos deductivas. Tienen la capacidad de deducir o inferir información a partir de datos almacenados.
- Bases de datos multimedia. Soportan video, música, libros electrónicos. Por lo regular son un conjunto de datos que abarcan un volumen muy importante. Su operación requiere un manejo especial para evitar la lentitud. Es importante sincronizar el audio y el video y mantener la calidad de las imágenes.
- Bases de datos temporales. Se distingue el tiempo de la validez, durante el cual un hecho se mantiene como verdadero, y el tiempo de transacción, que es el período en el cual se guarda un hecho en la base de datos.
- Bases de datos activas. Con un subsistema se gestionan definiciones y reglas. Se producen condiciones especiales para un conjunto de acciones.
- Bases de datos seguras. Tienen diferentes niveles de seguridad y confidencialidad. Contienen usuarios con distintos privilegios en materia de confidencialidad, accesibilidad e integridad de los archivos.
- Bases de datos difusas. Utilizan características del lenguaje natural (las palabras pueden significar diferentes cosas de acuerdo al contexto), por lo que este tipo de bases de datos se ajustan al modelo de inteligencia artificial.¹²

Extracción de datos

Murillo Danny, Saavedra Dalys, Quintero Erika, mencionan la búsqueda de este conocimiento en los datos que, con ayuda de la minería de datos, se basa en el uso de la metodología KDD (Knowledge Discovery in Databases) o Descubrimiento del Conocimiento a partir de Datos.¹³

¹² Gómez Fuentes, *Notas del curso de Bases de Datos*, 168-171.

¹³ Murillo Danny, Saavedra Dalys, Quintero Erika. Extracción de datos de perfiles en Google Scholar utilizando un algoritmo en el lenguaje R para hacer Minería de Datos, *Revista de I+D Tecnológico*, (2018): 95, <https://doi.org/10.33412/idt.v14.1.1807>. 261 Ídem.

Extraer conocimiento a partir de una base de datos es un cambio de paradigma que subraya la importancia del conocimiento útil a partir de los datos.

La minería de Texto identifica la información en la web, pero este conocimiento debe ser reducido a información útil.

Pasos para el procesamiento de datos

- Limpieza de datos. Consiste en llenar valores perdidos, disminuir el ruido y corregir inconsistencias.
- Integración de datos. Una integración cuidadosa de datos puede reducir redundancias e inconsistencias.
- Reducción de datos. Consiste en representar los datos en un volumen más pequeño.
- Transformación y discretización de datos. Es cambiar los datos para que tengan la forma apropiada para extraerse.¹⁴

Preparación de datos para IA

Sangüesa i Solé refiere que el objetivo principal de la preparación de datos es organizar los datos para que los programas de construcción de modelos puedan procesar estos datos y obtener el mejor modelo del conjunto de datos.

Dentro de las tareas del proceso de separación de datos están:

- Transformación de valores. Hacer que los datos puedan cambiar de valores sin perder información.
- Reducción de dimensionalidad. Consiste en eliminar casos y/o atributos con el fin de obtener modelos de calidad con menos recursos computacionales.¹⁵

¹⁴ Subirats Maté, Laia, Pérez Trenard, Diego Oswaldo, Calvo González, Mireia, Introducción a la limpieza y Análisis de los Datos, (Barcelona: Universidad Abierta de Cataluña, 2019),16-22.

¹⁵ Sangüesa i Solé, Ramón. Preparación de Datos (España: Universidad Abierta de Cataluña, 2019), p.7. 264 Serranos Castellanos, Manuel. La Protección Jurídica de las Bases de Datos, Revista Iberoamericana de Derecho Informático, No.8, (1995): 206.

III. Protección jurídica de las bases de datos

Convenio de Berna para la Protección de las Obras Literarias y Artísticas

Serranos Castellanos explica que el Convenio de Berna para la Protección de las Obras Literarias y Artísticas del 24 de julio de 1971 y enmendado el 28 de septiembre de 1979 protege a las bases de datos de acuerdo al texto de la citada convención siempre y cuando constituyan una creación original, no novedosa, es decir, represente una creación intelectual.

Acuerdo sobre los Aspectos de los Derechos de Propiedad Intelectual Relacionados con el Comercio

Artículo 10

2. Las compilaciones de datos o de otros materiales, en forma legible por máquina o de otra forma, que por razones de selección o de disposición de sus contenidos constituyan creaciones de carácter intelectual, serán protegidas como tales. Esta protección que no abarcará los datos o materiales en sí mismos. Se entenderá sin perjuicio de cualquier derecho de autor que subsista respecto de los datos o materiales en sí mismos.¹⁶

¿Qué dice la OMPI sobre las bases de datos?

En las sesiones de los comités de expertos celebradas en diciembre de 1994, la Delegación de la Comisión Europea informó a los comités sobre los progresos realizados en la labor de la Comunidad Europea en relación con la propuesta de la directiva sobre la protección jurídica que incluía una propuesta de creación de un derecho *sui generis* que se concedería al fabricante de una base de datos no original.¹⁷

¹⁶ Acuerdo sobre los Aspectos de los Derechos de Propiedad Intelectual Relacionados con el Comercio. Marrakech, 15 de abril de 1994.

¹⁷ Organización Mundial de la Propiedad Intelectual (OMPI), Propuesta Básica para las disposiciones sustantivas del Tratado relativo a la propiedad intelectual respecto de las Bases de Datos para consideración por la Conferencia Diplomática de Ginebra

Después, en las sesiones de los comités expertos celebradas en febrero de 1996, los comités y varias delegaciones expresaron un interés manifiesto en el derecho *sui generis* y en la continuación de la labor.

Producto de estas reuniones surge el proyecto de tratado relativo a la propiedad intelectual respecto a las bases de datos.

Redactó la propuesta básica para las disposiciones sustantivas del Tratado relativo a la propiedad intelectual, respecto de las bases de datos, para consideración por la Conferencia Diplomática de Ginebra de fecha 20 de diciembre de 1996.

Organización Mundial de la Propiedad Intelectual (OMPI), Propuesta Básica para las disposiciones sustantivas del Tratado relativo a la propiedad intelectual respecto de las bases de datos para consideración por la Conferencia Diplomática de Ginebra a 20 de diciembre de 1996

Preámbulo

Reconoce a las bases de datos como un elemento vital en el desarrollo de la infraestructura global de la información y un instrumento esencial para la protección del progreso económico, cultural y tecnológico.

Reconoce que la fabricación de bases de datos exige la inversión de recursos considerables, pero que dichas bases de datos pueden ser copiadas o es posible acceder a ellas a un costo inferior al costo necesario para diseñarlas en forma independiente.¹⁸

En su artículo 2, el Proyecto de tratado, fracción III, define al fabricante de la base de datos como la persona o personas naturales o jurídicas que tengan el control y la responsabilidad sobre una inversión sustancial en la fabricación de una base de datos.

El derecho y la protección estarán vinculados al empleador, sea este una empresa o cualquier otra organización que haga la inversión; la definición excluye a los subcontratistas que puedan haber sido contratados para dichas tareas.¹⁹

a 20 de diciembre de 1996 (Ginebra: 1996). 267 Ídem. 268 Ídem.

¹⁸ Ídem.

¹⁹ 271 ídem.

El término autor, fabricante de una base de datos, se aplica a los derechohabientes del fabricante de una base de datos.

El término inversión sustancial se define como cualquier inversión cualitativa o cuantitativamente importante de recursos humanos, financieros, técnicos o de otro tipo de recopilación, al ensamblaje, a la verificación, a la organización o a la presentación del contenido de la base de datos.

La Directiva 96/9/CE del Parlamento Europeo y del Consejo de 11 de marzo de 1996, sobre la protección jurídica de las bases de datos. Historia de la directiva

Rodríguez Manga menciona: En 1988, la Comisión de la Comunidad Europea presentó el documento titulado *Libro Verde* sobre los derechos de autor y el reto de la tecnología.²⁰

El 15 de abril de 1992 se presenta por la Comisión la Propuesta Originaria; es sometida al examen del Comité Económico y Social, que emite su dictamen el 24 de noviembre de 1992.

El 23 de junio de 1993, el Parlamento aprueba la propuesta de la Directiva presentada por la Comisión con un total de 32 enmiendas al texto.

La Comisión presenta una propuesta modificada el 4 de octubre de 1993. El 14 de diciembre de 1995, el Parlamento Europeo propuso nueve enmiendas al texto. El 26 de febrero de 1996 el Consejo aprueba las enmiendas y el 11 de marzo se adapta el texto como la Directiva 96/9/CE del Parlamento y del Consejo.²¹

Contenido de la Directiva

Artículo 5

Actos sujetos a restricciones

El autor de una base de datos tendrá el derecho exclusivo, respecto de la forma de expresión de dicha base susceptible de realizarse o autorizar a) la reproducción temporal o permanente, total o parcial; b) la traducción, adaptación, reordenación y cualquier otra

²⁰ Rodríguez Mangas Rubén, *La Protección Jurídica de las Bases de Datos*, Trabajo de fin de grado, Universidad de Salamanca, España, 2012. 273 Ídem.

²¹ Ídem.

modificación; c) cualquier forma de distribución al público de la base de datos o copias de la misma.

Excepciones a los actos sujetos a restricción

El usuario legítimo de una base de datos o de copias de la misma, podrá efectuar sin la autorización del autor de la base, todos los actos a que se refiere el artículo 5 que sean necesarios para el acceso al contenido de la base de datos y su normal utilización por el propio usuario.

En los siguientes casos, los Estados miembros podrán imponer limitaciones a los derechos contemplados en el artículo 5:

- a) Cuando se trate de una reproducción con fines privados de una base de datos no electrónica.
- b) Cuando la utilización se haga únicamente con fines ilustrativos o de enseñanza, o de investigación científica siempre que haga cita a la fuente.
- c) Cuando se utilice para fines de seguridad pública o a efectos de un procedimiento administrativo o judicial.
- d) Cuando se trate de otras excepciones a los derechos de autor tradicionalmente contempladas por su derecho interno”.²²

“Artículo 7. Derecho *Sui Generis*

Los Estados miembros dispondrán que el fabricante de la base de datos pueda prohibir la extracción y/o reutilización de la totalidad o de una parte sustancial del contenido de ésta, evaluada cualitativa o cuantitativamente, cuando la obtención, verificación o la presentación de dicho contenido representen una inversión sustancial desde el punto de vista cuantitativo o cualitativo”.²³

“Artículo 9. Excepciones al Derecho *Sui Generis*

Los Estados miembros podrán establecer que el usuario legítimo de una base de datos pueda, extraer y/o reutilizar una parte del contenido de la misma: a) cuando se trate de una extracción para fines privados; b) cuando se trate de una extracción con fines ilustrativos de enseñanza o de investigación científica citando la fuente y la medida está justificada por el objetivo no comercial que se persigue; c) cuando se trate de una extracción y/o reutilización para fines de seguridad pública o a efectos de un procedimiento administrativo o judicial.²⁴

²² Directiva 96/9/CE del Parlamento Europeo y del Consejo, de 11 de marzo de 1996, sobre la protección jurídica de las bases de datos.

²³ Ídem.

²⁴ Ídem.

Artículo 10. Plazo de la Protección

El derecho contemplado en el artículo 7 nacerá en el momento mismo en el que se dé por finalizado el proceso de fabricación de la base de datos. Expirará quince años después del 1 de enero del año siguiente a la fecha en que haya terminado dicho proceso. Cualquier modificación sustancial que conduzca a considerar que se trata de una nueva inversión sustancial permitirá atribuir a la base resultante de dicha inversión un plazo de protección propio.²⁵

Comité Permanente de Derechos de Autor y Derechos Conexos de la OMPI, Novena Sesión 23 a 27 de Junio de 2003

El Comité Permanente de Derechos de Autor y Derechos Conexos de la OMPI (SCCR) recibió una comunicación presentada por Kenya, que expresó una serie de cuestionamientos, entre los cuales estuvieron los siguientes:

“Los países en desarrollo no apoyaron la idea de crear un instrumento internacional para la protección de bases de datos no originales, aun cuando los distintos regímenes nacionales estaban en condiciones de crear la legislación para su protección”.²⁶

“Aunque se reconoce la inversión que se destina a la producción de los datos, las partes interesadas reconocen que algunas de las bases se compilan a partir de datos y obras protegidas por derechos de autor”.

Estudio sobre las limitaciones y excepciones relativas al derecho de autor y los derechos conexos en el entorno digital. Programas informáticos

Aunque el artículo 2.1 del Convenio de Berna no menciona expresamente a los programas informáticos, queda claro que en la mencionada convención existe la obligación de proteger esta categoría debido a su ubicación en la descripción general de las producciones en el campo literario, científico y artístico, cualquiera que sea el modo o forma de expresión.

²⁵ Ídem.

²⁶ 281 Ídem.

Compilaciones de datos o de otros materiales

El Artículo 10.1 de los Acuerdos sobre los Aspectos de los Derechos de Propiedad Intelectual relacionados con el Comercio (ADPIC) refiere a la obligación de proteger las compilaciones “como tales”. Es un artículo con una interpretación ambigua y, por tanto, los miembros de ese acuerdo tendrían la libertad de fijar el nivel de protección de esas compilaciones según sea conveniente para estos.²⁷

Comité Europeo de Protección de Datos, Dictamen 28/2024 sobre determinados aspectos de la protección de datos relacionados con el tratamiento de datos personales en el contexto de modelos de IA

Los modelos de IA, o simplemente modelos pertinentes para este dictamen, son los desarrollados a través del proceso de entrenamiento. En la fase de desarrollo en este proceso de entrenamiento, los modelos aprenden de los datos para realizar la tarea prevista. El proceso de entrenamiento requiere de un conjunto de datos a partir del cual el modelo identifica y aprende patrones. A través de diferentes técnicas, el modelo construirá una representación del conocimiento extraído del conjunto de datos de entrenamiento.²⁸

²⁷ Comité Permanente de Derechos de Autor y Derechos Conexos. Estudio sobre las limitaciones y excepciones relativas al derecho de autor y los derechos conexos en el entorno digital, Novena Sesión de 23 a 27 de junio de 2003, Organización Mundial de la Propiedad Intelectual (OMPI), 5 de abril 2003.

²⁸ Comité Europeo de Protección de Datos, Dictamen 28/2024 sobre determinados aspectos de la protección de datos relacionados con el tratamiento de datos personales en el contexto de modelos de IA, adoptado el 17 de diciembre de 2024.

Preparación y minimización de datos

El segundo ámbito de evaluación se refiere a la preparación de datos para la fase de entrenamiento. La autoridad de control deberá examinar en particular: I) si se ha considerado el uso de datos anónimos o personales que hayan sido objeto de seudonimización; II) en caso de que se haya decidido no utilizar tales medidas, los motivos de dicha decisión, teniendo en cuenta la finalidad prevista; III) las estrategias y técnicas de minimización de datos empleadas para restringir el volumen de datos personales, incluido el proceso de entrenamiento; IV) cualquier proceso de filtrado de datos aplicado al modelo antes de entrenarse con el fin de eliminar datos irrelevantes.²⁹

El tercer ámbito de evaluación se refiere a la selección de métodos en el desarrollo de modelos de IA

Las autoridades de control evaluarán las opciones para reducir o eliminar la identificabilidad en particular: I) si la metodología utiliza métodos para mejorar la generalización de los modelos, II) si el responsable aplicó técnicas adecuadas y eficaces para la conservación de la privacidad.

El último ámbito de evaluación se refiere a los métodos o las medidas que no afecten al riesgo de extracción directa de datos personales para el modelo por parte de cualquier persona que pueda acceder directamente a él, y que, además, pueda reducir la probabilidad de obtener datos personales relacionados con los datos de entrenamiento que se obtienen de las consultas.

IV. Propiedad Intelectual de Bases de datos en México

Los programas de cómputo y las bases de datos están protegidos en México por la Ley Federal de Derechos de Autor de acuerdo a los artículos que se mencionan en seguida:

Art. 101. Se entiende por programa de computación a la expresión original en cualquier lenguaje o código de un conjunto de instruc-

²⁹ Ídem. 285 Ídem.

ciones con una secuencia determinada que tiene como propósito que una computadora o dispositivo realice una tarea o función específica.

Art. 102. Los programas de computación se protegen en los mismos términos que las obras literarias.

Art. 107. Las bases de datos o de otros materiales legibles por medio de máquinas o en otra forma, que por razones de selección y disposición de su contenido constituyan creaciones intelectuales, quedarán protegidas como compilaciones. Dicha protección no se extenderá a los datos y materiales en sí mismos.

Art. 108. Las bases de datos que no sean originales, quedan protegidas en su uso exclusivo por quien las haya elaborado durante un lapso de 5 años.³⁰

En México se tiene como legislación de un derecho *sui generis* referente el artículo 108, que protege a las bases de datos no originales por un período de 5 años.

Parra Trujillo hace una mención referente: En la legislación mexicana, todas las bases de datos están protegidas por derechos de autor, mientras que en la directiva europea 96/9/CE las bases de datos no originales o que no sean fruto de una inversión sustancial carecen de protección.³¹

Aunque el artículo 108 es omiso en cuanto a si lo que está protegido en las bases de datos no originales es la estructura, el contenido o ambos.

Art. 109. El acceso a la información de carácter privado relativa a las personas contenidas en las bases de datos a que se refiere el artículo anterior, así como la publicación, reproducción, divulgación, comunicación pública y transmisión de dicha información, requerirá la autorización previa de las personas de que se trate. Quedan exceptuadas las investigaciones de la autoridad encargadas de la procuración e impartición de justicia, el acceso a los archivos públicos por las personas autorizadas por la ley siempre que la consulta sea realizada conforme a los procedimientos respectivos.

³⁰ Ley Federal de Derechos de Autor (México), Diario Oficial de la Federación, última reforma publicada el 1 de julio de 2020.

³¹ De la Parra Trujillo, Eduardo, El Derecho *Sui Generis* sobre las Bases de Datos en México y la Unión Europea, Revista Derecho Comparado de la Información, Número 3 (enero-junio 2004): 112, publicada por el Instituto de Investigaciones Jurídicas, UNAM, México.

Artículo 110. El titular del derecho patrimonial sobre una base de datos tiene el derecho exclusivo respecto a la forma de expresión de la estructura de dicha base, de autorizar o prohibir:

1. Su reproducción permanente o temporal, total o parcial, por cualquier medio y de cualquier forma.
2. Su traducción, adaptación, reordenación y cualquier otra modificación
3. La distribución del original o copias de la base de datos.
4. La comunicación al público.
5. La reproducción, distribución o comunicación pública de los resultados de las operaciones mencionadas en la fracción II del presente artículo.³²

La Ley Federal de Protección a la Propiedad Industrial en su artículo 5, expresa que el:

Instituto Mexicano de la Propiedad Industrial, la autoridad administrativa en materia de propiedad industrial, es un organismo descentralizado con personalidad jurídica, patrimonio propio el cual entre sus facultades serán: Fracción XXVIII. Promover la cooperación internacional mediante el intercambio de experiencias administrativas y jurídicas con instituciones encargadas del registro y protección legal de la propiedad industrial de otro país, incluyendo entre otras: la capacitación y el entrenamiento profesional de personal, la transferencia de metodologías de trabajo y la organización, el intercambio de publicaciones y la actualización de acervos documentales y bases de datos.³³

³² México, Ley Federal de Derechos de Autor, Diario Oficial de la Federación, última reforma publicada, el 1 de julio de 2020.

³³ México, Ley Federal de Protección a la Propiedad Industrial, Diario Oficial de la Federación. Publicada el 1 de julio de 2020.

V. La relación de las bases de datos para entrenamiento de las Redes Neuronales Recurrentes RNN y el TMEC

Artículo 20.8. Trato Nacional

1. Conforme a todas las categorías de propiedad intelectual cada parte otorgará a los connacionales de las otras partes un trato no menos favorable que a sus propios connacionales con relación a la protección de los derechos de propiedad intelectual. 2. La protección comprende los aspectos relativos al alcance, mantenimiento y observancia de los derechos de propiedad intelectual. La protección también incluye la prohibición de eludir las medidas tecnológicas efectivas (Medidas Tecnológicas de Protección. Y en las disposiciones a la formación sobre la gestión de derechos. Artículo 20.67. Medidas tecnológicas de Protección.

6. Medida tecnológica efectiva es cualquier dispositivo, tecnología o componente que controla el acceso a una obra, interpretación o ejecución o fonograma protegidos por derechos de autor o un derecho conexo.

Artículo 20.68. Información sobre la Gestión de Derechos.

Con la finalidad de proporcionar recursos legales adecuados y efectivos para proteger la información sobre la gestión de derechos (IGD), cada parte firmante establecerá que cualquier persona que, sin autorización, y a sabiendas, o teniendo motivos razonables para saber, que podría facilitar, permitir, inducir o encubrir una infracción al derecho de autor o derechos conexos de autores, intérpretes o productores de fonogramas, será legalmente responsable si:

a) Elimina o modifica cualquier elemento de la gestión de derechos, b) distribuya o importe para su distribución la gestión de derechos IGD sabiendo que ha sido alterada sin tener la autorización respectiva, c) distribuya, importe para su distribución, transmita, comunique o ponga a disposición del público copias de las obras, interpretaciones o ejecuciones de fonogramas sabiendo que la gestión de derechos IGD ha sido suprimida o alterada sin la autorización respectiva.

70. Una medida tecnológica que de manera usual puede evitarse por accidente no es una medida tecnológica efectiva.

73. Una parte podrá cumplir con sus obligaciones si proporciona protección a las compilaciones originales siempre que los actos

descritos en este subpárrafo se traten como derechos de autor en tales compilaciones originales.³⁴

Aunque el T-MEC no contempla un derecho *sui generis* para las bases de datos no originales, el artículo 20.2 hace referencia a que los derechos de propiedad intelectual deben contribuir a la promoción de la innovación tecnológica y a la transferencia y difusión de la tecnología. El beneficio debe ser recíproco entre los productores y usuarios del conocimiento, de modo que favorezca el bienestar social y económico y un equilibrio entre derechos y obligaciones.

También el artículo 20.3 sobre los principios establece que una parte contratante, al modificar o formular su regulación, podrá adoptar las medidas necesarias para prevenir el abuso al derecho de propiedad intelectual o las prácticas que limiten de manera injustificable el comercio o afecten de forma negativa la transferencia internacional de tecnología.³⁵

El artículo anterior está relacionado con las inversiones sustanciales de las que habla la Directiva Europea 96/9/CE sobre la protección jurídica de las bases de datos.

VI. Conclusiones

Un argumento para la protección jurídica de las bases de datos, para el entrenamiento de las redes neuronales artificiales, será el concepto de inversión sustancial, el cual vincula los derechos con el empleador y que excluye a los subcontratistas con acceso a esos datos y tareas.

De acuerdo al dictamen europeo 28/2024, sería posible extraer y utilizar los datos a una base de datos no original, si los datos están anonimizados o pseudonimizados para reducir su identificabilidad.

En México se cuenta con la protección a las bases de datos no originales, pero no hace referencia al requisito de la inversión sustancial del que sí menciona la Directiva 96/9/CE del Parlamento Europeo.

Canadá no contempla un derecho *sui generis* para la protección de bases de datos para el entrenamiento de IA. Algunos conceptos son apli-

³⁴ Tratado entre los Estados Unidos Mexicanos, los Estados Unidos de América y Canadá.

³⁵ México, Tratado entre México, Estados Unidos y Canadá T-MEC, Diario Oficial de la Federación. 29 de junio de 2020.

cables a dicha figura, como el de trato no menos favorable que a sus connacionales, o eliminar las prácticas que afecten de forma injustificable la transferencia internacional de tecnología.

Sigue el debate abierto sobre si lo que se protege es el contenido de las bases, o si, aunque el contenido de las bases no tenga protección, lo que se protege es la inversión sustancial.

Los modelos de IA seguirán necesitando de grandes bases de datos para su entrenamiento y la protección jurídica será un elemento sustancial para el desarrollo o mejoramiento de nuevos modelos de inteligencia artificial.

VII. Bibliografía

Basogain, X. (2008). Redes neuronales artificiales y sus aplicaciones [Material de curso]. Open Course Ware, Universidad del País Vasco. <http://ocw.ehu.eus/enseñanzas-tecnicas/redes-neuronales-artificiales-y-sus-aplicaciones/course-listing>

Comité Europeo de Protección de Datos. (17 de diciembre de 2024). Dictamen 28/2024 sobre determinados aspectos de la protección de datos relacionados con el tratamiento de datos personales en el contexto de modelos de IA.

Comité Europeo de Protección de Datos. (17 de diciembre de 2024). Dictamen 28/2024 sobre determinados aspectos de la protección de datos relacionados con el tratamiento de datos personales en el contexto de modelos de IA. https://www.edpb.europa.eu/our-work-tools/our-documents/opinion-board-art-64/opinion-282024-certain-data-protection-aspects_es

Comité Permanente de Derechos de Autor y Derechos Conexos, Novena Sesión de 23 a 27 de Junio de 2003. Comunicación presentada por Kenya. Organización Mundial de la Propiedad Intelectual (OMPI). (Documento SCCR/9/7) https://www.wipo.int/edocs/mdocs/copyright/es/sccr_9/sccr_9_11.pdf

Date, C. J. (2001). Introducción a los sistemas de bases de datos. Prentice Hall.

De la Parra Trujillo, E. (2004). El derecho sui generis sobre las bases de datos en México y la Unión Europea. Revista Derecho Comparado de

- la Información, (3), 101–124. Instituto de Investigaciones Jurídicas, Universidad Nacional Autónoma de México.
- Estados Unidos Mexicanos, Estados Unidos de América, y Canadá. (2020, 29 de junio). Tratado entre México, Estados Unidos de América y Canadá (T-MEC). Diario Oficial de la Federación.
- Estados Unidos Mexicanos. (2020). Ley Federal de Derechos de Autor. Diario Oficial de la Federación.
- Estados Unidos Mexicanos. (2020). Ley Federal de Protección a la Propiedad Industrial. Diario Oficial de la Federación
- Fernández Iglesias, J. M. (2023). Pequeña introducción a las bases de datos [Informe técnico]. Universidad de Vigo. <https://doi.org/10.13140/RG.2.2.13924.65925>
- Gómez Fuentes, M. del C. (2013). Notas del curso bases de datos [Apuntes de curso]. Universidad Autónoma Metropolitana.
- Isasi Viñuela, P., y Galván León, I. (2004). Redes neuronales artificiales: Un enfoque práctico. Pearson.
- Moreno Pallares, M., y Moreno Pallares, R. (2022). Entrenamiento de redes neuronales artificiales con aprendizaje no supervisado en el control de calidad del análisis. *Polo del Conocimiento*, 7(5), 1584–1593. <https://doi.org/10.23857/pc.v7i5.4048>.
- Murillo, D., Saavedra, D., y Quintero, E. (2018). Extracción de datos de perfiles en Google Scholar utilizando un algoritmo en el lenguaje R para hacer minería de datos. *Revista de I+D Tecnológico*, 14(1), 94–104. <https://doi.org/10.33412/idt.v14.1.1807>
- Organización Mundial de la Propiedad Intelectual. (1971). Convenio de Berna para la Protección de las Obras Literarias y Artísticas. https://www.wipo.int/edocs/pubdocs/es/wipo_pub_287-accessible1.pdf
- Organización Mundial de la Propiedad Intelectual. (20 de diciembre de 1996). Propuesta básica para las disposiciones sustantivas del tratado relativo a la propiedad intelectual respecto de las bases de datos para consideración por la Conferencia Diplomática.
- Organización Mundial del Comercio. (1994). Acuerdo sobre los Aspectos de los Derechos de Propiedad Intelectual Relacionados con el Comercio (ADPIC).
- Parlamento Europeo y Consejo de la Unión Europea. (1996, 11 de marzo). Directiva 96/9/CE sobre la protección jurídica de las bases de datos.

- Diario Oficial de las Comunidades Europeas. <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:31996L0009>
- Quezada Peralta, G., Donate López, J., y Abreu González, R. (s.f.). Análisis comparativo de los grandes modelos de lenguaje. Monografías Secoir. <https://secoir.org/inteligencia-artificial-generativa/>
- Raraz Vidal, J. (2023). La importancia de las bases de datos para el entrenamiento en inteligencia artificial. *Revista de Investigación en Salud*, 7(3), 121–122. <https://doi.org/10.35839/repis.7.3.1970>
- Rodríguez Mangas, R. (2012). La protección jurídica de las bases de datos. Universidad de Salamanca.
- Sangüesa i Solé, R. (2019). Preparación de datos [Material de curso]. Universidad Abierta de Cataluña.
- Schiaffino, S. (2018). Inteligencia artificial: Técnicas de clasificación [Documento de trabajo]. Universidad Nacional del Centro de la Provincia de Buenos Aires. https://users.exa.unicen.edu.ar/catedras/optia/public_html/clasificacion.pdf
- Serranos Castellanos, M. (1995). La protección jurídica de las bases de datos. *Revista Iberoamericana de Derecho Informático*, (8), 205–210.
- Sierra Ramos, J. M., y López Montes, A. (2022). Introducción a las redes neuronales artificiales [Material de curso]. Universidad Complutense de Madrid.
- Subirats Maté, L., Pérez Trenard, D. O., y Calvo González, M. (2019). Introducción a la limpieza y análisis de los datos. Universidad Abierta de Cataluña.

