



Disponible en línea en www.sciencedirect.com

ScienceDirect

página de inicio de la revista: www.elsevier.com/locate/CLSR

**Computer Law
&
Security Review**



Ajuste fino de GPT-3 para la clasificación de reglas legales

Davide Ligaa, ^a, Livio Robaldo b.

^a un grupo de razonamiento individual y colectivo (ICR), Universidad de Luxemburgo, Esch-sur-Alzette, Luxemburgo

^b Laboratorio de Innovación Legal de Gales, Universidad de Swansea, Singleton Park, Sketty, Swansea, Reino Unido

información del artículo

Palabras clave:

Clasificación de reglas

GPT-3

IA y derecho

abstracto

En este artículo, proponemos una tarea de Clasificación de reglas legales (LRC) utilizando uno de los modelos de lenguaje más discutidos en el campo de la Inteligencia Artificial, a saber, GPT-3, un modelo de lenguaje generativo previamente entrenado. Entrenamos y probamos la tarea LRC propuesta sobre el GDPR codificado en LegalDocML (Palmirani y Vitali, 2011) y LegalRuleML (Athán et al., 2013), dos estándares XML ampliamente utilizados para el dominio legal. Utilizamos las anotaciones LegalDocML y LegalRuleML proporcionadas en Robaldo et al. (2020) para perfeccionar GPT-3. Si bien mostramos la capacidad de los modelos de lenguaje grandes (LLM) para aprender fácilmente a clasificar reglas legales y deonticas incluso en pequeñas cantidades de datos, mostramos que GPT-3 puede superar significativamente experimentos anteriores en la misma tarea. Nuestro trabajo se centró en una tarea multiclase, demostrando que GPT-3 es capaz de reconocer la diferencia entre reglas de obligación, reglas de permiso y reglas constitutivas con desempeños que superan puntuaciones anteriores en LRC.

© 2023 Davide Liga y Livio Robaldo. Publicado por Elsevier Ltd. Todos los derechos reservados.

1. Introducción

Los recientes desarrollos en Inteligencia Artificial (IA), en el campo del Procesamiento del Lenguaje Natural (PLN), así como en otros campos como la Visión por Computador y el Reconocimiento de Voz, han demostrado el poder innovador de los Modelos de Lenguaje Grande (LLM) previamente entrenados.

El éxito de los LLM comenzó casi en 2018, con BERT de Google (Devlin et al.) y GPT de OpenAI (Radford y Narasimhan, 2018). La idea detrás de los LLM es sencilla: podemos crear modelos de lenguaje potentes entrenando arquitecturas neuronales en grandes cantidades de datos. En este sentido, la arquitectura neuronal de la que derivan muchos LLM es la arquitectura transformadora (Vaswani et al., 2017). Estos modelos de lenguaje han demostrado la capacidad de aplicarse a una amplia gama de tareas de PNL, superando el estado del arte en muchos desafíos de PNL.

En este artículo, nos centramos en la Clasificación de Reglas Jurídicas (LRC), que ha sido abordada solo por unos pocos académicos en el campo de la Inteligencia Artificial y el Derecho (AI&Law), a pesar de su importancia (Robaldo et al., 2019).

Comparamos el desempeño de los pocos estudios previos que emplean LLM con el desempeño de nuestro trabajo, que emplea dos versiones mejoradas de GPT-3. Hasta donde sabemos, este artículo es el primer intento de perfeccionar el conocido GPT-3 de OpenAI (actualmente, el modelo de lenguaje más poderoso del mundo) en datos legales y, en particular, en el reconocimiento de reglas deonticas.

La capacidad de detectar automáticamente reglas deonticas a partir del lenguaje natural es una dirección de investigación importante para toda la comunidad de AI&Law. Permitirá el desarrollo de sistemas expertos jurídicos avanzados (cf. (Boella et al., 2016)); Además, el uso de LLM fomentará específicamente una integración más profunda de la IA basada en datos "de abajo hacia arriba" con la IA simbólica "de arriba hacia abajo", es decir, una integración de los LLM con las teorías teóricas más recientes.

Autor de correspondencia: Davide Liga, Grupo de Razonamiento Individual y Colectivo (ICR), Universidad de Luxemburgo, 6 av. de la Fonte, 4364, Esch-sur-Alzette, Luxemburgo Dirección de correo electrónico: davide.liga@uni.lu (D.Liga).

<https://doi.org/10.1016/j.clsr.2023.105864>

0267-3649/© 2023 Davide Liga y Livio Robaldo. Publicado por Elsevier Ltd. Todos los derechos reservados.

resulta en lógica y argumentación deónticas formales (Wyner y Pedro, 2011; Ashley, 2017; Sol y Robaldo, 2017). En este trabajo, mostramos:

- Cómo GPT-3 puede lograr resultados notables en el reconocimiento reglas legales y deónticas
- Cómo se puede emplear el conocimiento simbólico existente para explotar aún más el potencial de GPT-3

Aunque la idea de utilizar IA para extraer automáticamente reglas y modalidades deónticas del texto legal no es nueva (ver Sección 2), existen algunos obstáculos importantes que normalmente impidió que la comunidad AI&Law lograra mejores resultados, es decir, la falta de datos disponibles diseñados ad hoc para el clasificación de reglas y modalidades deónticas. De hecho, anotar y crear este tipo de datos no sólo requiere mucho tiempo. pero también costoso y requiere expertos en el campo, que no son siempre disponible.

Por otro lado, la creación de conjuntos de datos de alta calidad. suele estar en equilibrio con el tamaño de estos conjuntos de datos, que a su vez limitan la precisión de los clasificadores estándar de aprendizaje automático (por ejemplo, (Boella et al., 2013)), especialmente aquellos que emplean arquitecturas neuronales profundas (por ejemplo, (Song et al., 2022)), que notoriamente necesitan grandes cantidades de datos.

En este sentido, los LLM recientemente allanaron el camino hacia una nueva paradigma en IA, a veces denominado "Aprendizaje por transferencia", lo que indica la idea de que podemos usar LLM transfiriendo lo que "aprendieron" durante su fase previa a la capacitación para tareas posteriores y datos posteriores. Esto demostró la capacidad de estos modelos para lograr resultados impresionantes incluso en pequeños conjuntos de datos. Por estas razones, muchos investigadores comenzaron a utilizar LLM como BERT (Devlin et al.), que es uno de los más famosos ejemplos de arquitecturas neuronales previamente entrenadas exitosas, utilizadas en muchas tareas posteriores (incluidas tareas que empleaban conjuntos de datos muy pequeños (Liga y Palmirani, 2020)).

En los últimos años, algunos estándares XML legales han sido propuestos, que son capaces de proporcionar a los investigadores conocimiento jurídico legible por máquina. El más popular es sin duda Akoma Ntoso, también conocido como LegalDocML1, que se utiliza representar documentos legales en formato XML, permitiendo así codificar la estructura de los documentos legales (secciones, preámbulos, artículos, etc.), así como metadatos relacionados con la naturaleza y la historia de dichos documentos. Otro estándar XML legal es LegalRuleML2, que es capaz de representar la lógica dimensión de los documentos legales, incluidas las reglas deónticas lógicas, relacionados, por ejemplo, con obligaciones y permisos.

En este trabajo, queremos utilizar estos estándares XML en combinación con uno de los más PML del mundo, y probablemente el más popular hasta ahora, a saber, GPT-3. Específicamente, este artículo mostrará el potencial del uso de documentos XML legales como fuente de datos para aplicar GPT-3 en tareas posteriores como como LRC.

Como se dijo anteriormente, LRC, y en particular la clasificación automática de reglas deónticas del lenguaje natural, es una tarea que ha recibido poca atención por parte de la comunidad AI&Law, a pesar de su utilidad dentro de los sistemas expertos legales. Esta tarea consiste en clasificar sentencias jurídicas únicas o proposiciones jurídicas únicas.

visiones que contienen modalidades deónticas tales como Obligaciones, Prohibiciones y permisos (Nguyen et al., 2022; Liga y Palmirani, 2022a).

En la Sección 2, describiremos algunos trabajos relacionados, mientras que La Sección 3 describirá brevemente el papel de los estándares LegalXML. En la Sección 4, brindaremos una breve descripción general de nuestro método, incluida una breve introducción tanto a la técnica de extracción de datos (2.1) como a la técnica de clasificación (2.2). En las siguientes dos secciones, daremos una descripción más exhaustiva de los datos recuperados (Sección 5), y un informe detallado sobre los entornos experimentales con sus respectivos resultados (Sección 6). En particular, presentamos cuatro escenarios experimentales y dos estrategias rápidas, es decir, ocho escenarios en total. Finalmente, La sección 7 concluye el artículo.

2. Obras relacionadas

Ha habido (relativamente pocos) intentos en la literatura de emplear metodologías de PNL para detectar automáticamente reglas en el ámbito jurídico. documentos. Entre estos primeros intentos se encuentran estudios abordar la clasificación de elementos deónticos (Dragoni et al., 2018) como parte de una gama más amplia de objetivos (Kiyavitskaya et al., 2008; Waltl et al., 2017; Gao y Singh, 2014; de Maat y Winkels, 2008). Entre estos primeros intentos de clasificar las obligaciones a partir de textos legales se encuentra Kiyavitskaya et al. (2008), que se centró en las regulaciones de Italia y Estados Unidos. Su El método empleó listas de palabras, gramáticas y heurísticas para extraer obligaciones entre otros objetivos como derechos y restricciones. Otro trabajo que abordó la clasificación de declaraciones deónticas es Waltl et al. (2017), que se centró en la Ley de arrendamiento alemana y 22 clases de declaraciones clasificadas (entre los cuales también hubo prohibiciones y permisos). El método utilizado es aprendizaje activo con ingenio multinomial. bayes, regresión logística y clasificadores de perceptrones multicapa, sobre un corpus de 504 frases. En Gao y Singh (2014), Los autores utilizaron el aprendizaje automático para extraer seis clases. de relaciones normativas: prohibiciones, autorizaciones, sanciones, compromisos y poderes.

Como se mencionó anteriormente, los primeros estudios que abordaron la clasificación de reglas deónticas centradas en elementos deónticos como partes de una gama más amplia de objetivos. Quizás el primer estudio que Centrado principalmente en la esfera deóntica es Neill et al. (2017). Este El trabajo se centró en la legislación financiera para clasificar sentencias legales utilizando una arquitectura Bi-LSTM, con una capacitación conjunto de datos que contiene 1297 instancias (596 obligaciones, 94 prohibiciones y 607 permisos).

Si bien la mayoría de los estudios relacionados con LRC son Inteligencia Artificial fuertemente enfocada simbólica y basada en reglas (Wyner y Peters, 2011, de Maat y Winkels, 2008), algunos estudios se centraron en cambio en el uso de redes rurales (Neill et al., 2017, Chalkidis et al.) y LLM (Shaghaghian et al., 2020; Joshi et al., 2021; Liga y Palmirani, 2022a). Hasta donde sabemos, el primer estudio que empleó un modelo de lenguaje (es decir, BERT) para la clasificación de oraciones deónticas son Shaghaghian et al. (2020) y Joshi et al. (2021). Shaghaghian et al. (2020) utilizaron cuatro pre-entrenados arquitecturas (BERT, DistilBERT, RoBERTa y ALBERT) pero centrado solo en las tareas de detección binaria frente a las no tareas. Joshi y cols. (2021) también se centró en los permisos, logrando un

¹ <https://www.oasis-open.org/committees/legaldocml>.

² <https://www.oasis-open.org/committees/legalruleml>.

precisión promedio y recuperación del 90% y 89,66% respectivamente. A Un trabajo reciente, en el que se inspira este artículo, mostró cómo utilizar BERT ([Devin et al.](#)), DistilBERT ([Sanh et al.](#)) y Legal-BERT ([Chalkidis et al.](#)) para clasificar obligaciones, permisos y reglas constitutivas automáticamente ([Liga y Palmirani, 2022a](#)), inspirado a su vez en resultados positivos anteriores utilizando Tree Kernel algoritmos ([Liga y Palmirani, 2022b](#)).

Entre estos estudios, ninguno intentó utilizar la tecnología generativa. LLM, que recientemente obtuvo un inmenso éxito gracias a la lanzamiento de GPT-3 y ChatGPT ([Brown et al., 2020](#)). En este trabajo, Queremos cubrir esta brecha, fomentada también por el hecho de que GPT-3 es actualmente el modelo de lenguaje más poderoso, o al menos el uno que recibió la atención más entusiasta por parte de los medios y comunidades científicas.

Por lo tanto, al igual que [Liga y Palmirani \(2022a\)](#), nuestro trabajo presenta un enfoque de aprendizaje por transferencia de aprendizaje automático, que combina la información simbólica de los formatos XML legales con el poder subsimbólico proporcionado por GPT-3. Por Aprovechando la información canalizada por la mayor base de conocimiento de Legal-RuleML disponible, presentamos cuatro diferentes escenarios de clasificación:

1. Regla versus no regla
2. Deóntico versus no deóntico
3. Obligación versus permiso versus no deóntico
4. Obligación versus permiso versus regla constitutiva versus no regla

La novedad y el poder de las metodologías de IA generativa junto con el uso combinado de LegalDocML y Legal-RuleML son dos contribuciones importantes de este estudio, junto con el diseño de los entornos experimentales en cuatro escenarios de clasificación diferentes utilizando dos estrategias de estímulo diferentes. En Además, los formatos XML legales como LegalDocML y Legal-RuleML suelen estar escritos y validados por expertos legales. En En otras palabras, para la tarea de detectar clases deónticas, la extracción de datos de este tipo de documentos podría decirse que puede ofrecer una solución más conveniente y robusta en comparación con la uso de conjuntos de datos de propósito general que podrían ser sólo parcialmente relacionado con la esfera deóntica.

3. XML legal

Este trabajo explota el potencial de dos reconocidos LegalXML estándares: LegalDocML y LegalRuleML, que son claves en el campo de AI&Law y Representación del Conocimiento Jurídico.

LegalDocML aplica el estándar XML específicamente a documentos legales. Este sistema codifica documentos legales y legislativos en un formato estructurado y legible por máquina, facilitando su gestión e intercambio entre diversos sistemas y plataformas. LegalDocML es muy valorado por su capacidad de agilizar el procesamiento y análisis de los asuntos legales documentos. Esto se logra estandarizando el formato de estos documentos, permitiendo su interpretación, visualización, y se almacena de manera uniforme en cualquier sistema. Además, Legal-DocML refuerza la precisión y eficiencia de las comparaciones de documentos legales y mejora su interoperabilidad. En consecuencia, esto facilita a los profesionales del derecho, las instituciones y a los desarrolladores de software compartir y utilizar información armoniosamente en un formato estandarizado.

LegalRuleML abarca características para representar el normas y reglas en documentos legales. Al igual que LegalDocML, LegalRuleML también está basado en XML, lo que permite la incorporación de reglas y lógica en un formato legible por máquina. LegalRuleML es ventajoso ya que permite el razonamiento automatizado sobre el contenido de los documentos legales a partir de disposiciones normativas perspectiva. Ayuda a gestionar y manipular grandes volúmenes de normas jurídicas, vinculando reglas con documentos legales o legislativos relevantes basados en teorías jurídicas subyacentes. Esencialmente, LegalRuleML presenta la dimensión normativa de lo legal. documentos computacionalmente comprensibles, mejorando así la accesibilidad y facilitando el desarrollo de tecnologías avanzadas. Soluciones tecnológicas legales. Estos incluyen la automatización de documentos, extracción de conocimientos y análisis de textos jurídicos. A través de esto, Los sistemas pueden garantizar una interpretación y aplicación consistentes de las leyes, resoluciones legales y políticas, impulsando así la eficiencia y precisión de los sistemas legales automatizados.

4. Metodología

El objetivo de este trabajo es doble. Por un lado, tiene como objetivo mostrar que LegalDocML y LegalRuleML pueden ser combinados para alimentar algoritmos de aprendizaje automático de IA generativa con datos confiables para la clasificación de reglas deónticas. En por otro lado, pretende probar el uso del aprendizaje por transferencia en la tarea de clasificación de reglas deónticas. El primer aspecto (es decir, la combinación de LegalDocML y LegalRuleML) está relacionada con la metodología que se ha utilizado para extraer el conocimientos y datos jurídicos. El segundo aspecto (es decir, la era de uso del aprendizaje por transferencia como algoritmo de aprendizaje automático) está relacionado con la metodología de clasificación. La combinación de estos dos aspectos metodológicos (es decir, método de extracción de datos/conocimiento y método de clasificación) definirse como un enfoque de IA híbrida, ya que combina simbólicos conocimiento con conocimiento subsimbólico ([Gómez-Pérez et al., 2020](#), [Rodríguez-Doncel et al., 2020](#), [Liga](#)).

4.1. Método de extracción de datos

En cuanto a la metodología, replicamos el enfoque propuesto en [Liga y Palmirani \(2022b\)](#) y [Liga y Palmirani \(2022a\)](#), donde se utilizaron algunos clasificadores de aprendizaje automático. capacitado, en entornos supervisados, sobre la información contenida en dos formatos XML legales, a saber, LegalDocML y LegalRuleML.

Al igual que en [Liga y Palmirani \(2022\)](#), un importante aporte de Este trabajo trata sobre la novedad y el poder del aprendizaje por transferencia. metodologías conjuntamente con el uso combinado de LegalDocML y LegalRuleML3.

De manera similar a [Liga y Palmirani \(2022b\)](#) y [Liga y Palmirani \(2022a\)](#), este trabajo utilizó un conjunto de datos de 707 disposiciones legales atómicas que se encuentran en el Reglamento General Europeo de Protección de Datos (GDPR). Los datos fueron obtenidos del DAPRECO. Knowledge Base (D-KB para abreviar) ([Robaldo et al., 2020](#)), que es la base de conocimientos más grande en LegalRuleML, así como la

³ Además, formatos XML legales como LegalDocML y LegalRuleML. son documentos escritos por expertos legales, que proporcionan a los algoritmos de aprendizaje automático datos de alta calidad.

Tabla 1 – Número de instancias por clase por escenario.

	Clases	Instancias		Clases	Instancias
escenario 1	Regla	260	Escenario 2	deóntico	204
	no regla	447		no deóntico	503
Escenario 3	Clases	Instancias	Escenario 4	Clases	Instancias
	obligación	156		obligación	156
	permiso no	44		permiso	44
	deóntico	503		constitutivo no	56
				regla	447

base de conocimiento más grande formalizada en Entrada/Salida cosificada Lógica (Robaldo, 2021).

El D-KB representa la dimensión lógico-deóntica de la GDPR y actualmente contiene 966 cosificados. fórmulas lógicas de E/S, incluyendo 271 obligaciones, 76 permisos y 619 constitutivos. normas. Las reglas constitutivas se utilizan para desencadenar inferencias específicas. para las reglas modeladas y se distinguen de las obligaciones o permisos en que no transmiten información sobre deóntic modalidades. DAPRECO también incluye conexiones entre cada fórmula y su correspondiente parte estructural de la ley documento en la representación LegalDocML del GDPR, permitiendo un vínculo entre los aspectos lógico-deónticos de los documentos legales, representados por las fórmulas 966 Entrada/Salida en DAPRECO, y las declaraciones en lenguaje natural en el texto legal, representado por la representación LegalDocML de el GDPR. La combinación de LegalDocML y LegalRuleML es También es útil para reconstruir el objetivo exacto del lenguaje natural, como por ejemplo combinando listas de obligaciones en textos legales en una sola frase. En este sentido, el papel de LegalDocML es particularmente importante, ya que permite la reconstrucción de los sistemas atómicos. disposiciones (por ejemplo, cuando una disposición se divide en una lista).

Combinando la información estructural de LegalDocML y la información deóntica en LegalRuleML extrajimos 707 disposiciones legales etiquetadas en total. Las etiquetas de estas oraciones. son los mismos que los proporcionados por DAPRECO con la adición de una categoría "sin regla". Abreviamos "obligationRule", "permissionRule", "constitutiveRule" en "obligación", "permis-sion" y "constitutive" respectivamente.

La clase "obligación" se refiere a aquellas oraciones que tener al menos una regla de obligación en sus fórmulas relacionadas. El La clase "permiso" se refiere a aquellas oraciones que tienen al menos una regla de permiso en sus fórmulas relacionadas. La clase "Constitutivas" se refiere a aquellas oraciones que simplemente son reglas constitutivas en sus fórmulas relacionadas. También consideramos un clase "no regla", que se refiere a todas las oraciones que tienen ninguna regla legal en absoluto, y "no deóntico", que se refiere a todos sentencias que no contienen obligaciones ni permisos (aunque pueden contener reglas constitutivas)⁴.

Estas etiquetas permiten cuatro configuraciones experimentales diferentes, como se muestra en la Tabla 1, que proporcionan diferentes niveles de granularidad como se muestra en la Fig. 2.

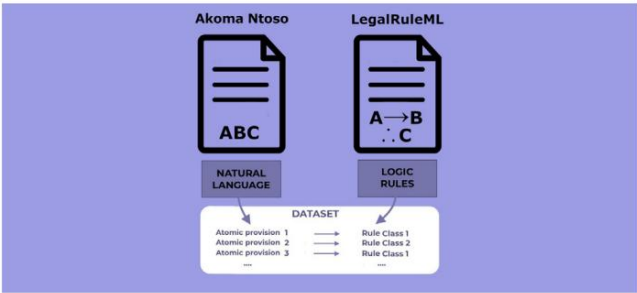


Fig. 1 – Extracción de conocimiento de LegalDocML y LegalRuleML. Tenga en cuenta que cada instancia extraída se refiere a una disposición normativa atómica (generalmente contenida en párrafos o puntos), y a veces pueden consistir en más de una frase.

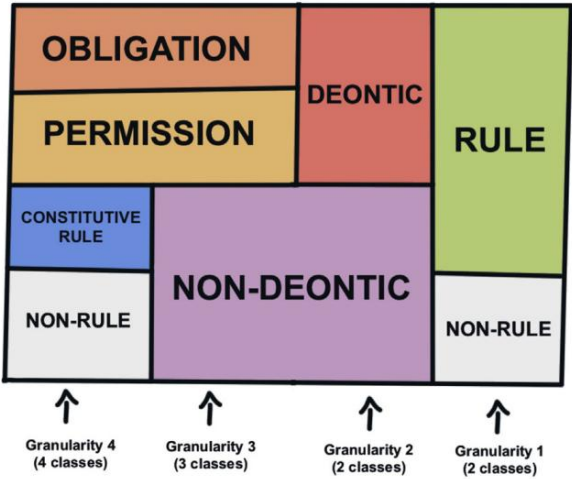


Fig. 2 – Niveles de granularidad de las cuatro clasificaciones escenarios.

4.2. Método de clasificación

En cuanto al método de clasificación del aprendizaje por transferencia, GPT-3 afinado (Brown et al., 2020) en nuestro conjunto de datos de 707 disposiciones legales extraídas. Para el enfoque de ajuste, diseñamos dos indicaciones simples diferentes para darle a GPT-3 las instrucciones necesarias para clasificar las clases de provisiones atómicas.

⁴ Para las clasificaciones múltiples (es decir, Escenario 3 y 4), se eliminaron cuatro afirmaciones, ya que las clases "obligación" y "permisión" se superponían.

```
{"prompt":"The data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the personal data and the following information the recipients or categories of recipient to whom the personal data have been or will be disclosed, in particular recipients in third countries or international organisations;->", "completion":" 1"}"
```

Fig. 3 – Ejemplo de disposición legal con el mensaje 1.

```
{"prompt":"The data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the personal data and the following information the recipients or categories of recipient to whom the personal data have been or will be disclosed, in particular recipients in third countries or international organisations;\n\nThe previous text is a ->", "completion":" obligation"}"
```

Fig. 4 – Ejemplo de disposición legal con el mensaje 2.

Nuestro primer mensaje tiene la siguiente plantilla:
mensaje: "[DISPOSICIÓN LEGAL ATÓMICA]->"
finalización: "[CLASE COMO NÚMERO]"

Donde "[DISPOSICIÓN LEGAL ATÓMICA]" es la única atómica disposición extraída de LegalDocML y LegalRuleML, mientras que "->" es nuestro marcador de clasificación, mientras que "\n" representa una nueva línea. La finalización de este mensaje es la clase representada como un número (en el caso del escenario 4, los números 0, 1, 2, 3 significan "ninguno", "obligación", "permiso" y "constitutivo" respectivamente).

También probamos un segundo mensaje, a saber:
mensaje: "[DISPOSICIÓN LEGAL ATÓMICA]\n\nLa el texto anterior es un ->"
finalización: esta "[NOMBRE DE LA CLASE]"
vez, la finalización de este mensaje es la clase del disposición legal atómica representada como palabras (no como números). Un ejemplo de disposición legal marcada usando el mensaje 1 puede ser visto en la Fig. 3.

Un ejemplo de disposición legal marcada usando el mensaje 2 puede verse en la Fig. 4.

mientras que las reglas constitutivas son un indicador de la existencia de una Por regla general, no dan información sobre las modalidades deónticas.

Un aspecto importante del D-KB es que contiene la referencias a los elementos estructurales (párrafos, puntos, etc.) donde cada norma del GDPR, representada en el D-KB como Se puede encontrar una regla si-entonces. Más precisamente, se refiere a la Representación LegalDocML del GDPR, donde todos los estructurales Los elementos del RGPD se pueden encontrar siguiendo LegalDocML. estándares⁷. En otras palabras, utilizar un conocimiento de LegalRuleML base como el D-KB y la correspondiente representación LegalDocML, es posible conectar la esfera lógico-deóntica de documentos legales a las declaraciones en lenguaje natural en el texto legal, proporcionado por la representación LegalDocML de la RGPD (ver Figs. 1 y 5).

Es importante señalar que esta combinación de LegalDocML y LegalRuleML también facilitan la reconstrucción del objetivo exacto, en términos de lenguaje natural, donde se ubica cada disposición. Por ejemplo, muchas obligaciones de Los textos legales se dividen en listas, y LegalDocML es crucial para reconstruir esas piezas de lenguaje natural en una pieza única. del lenguaje natural. Por ejemplo, el artículo 5 del RGPD8 establece:

5. Datos

En este estudio, empleamos un conjunto de datos que contiene una cantidad total de 707 disposiciones normativas atómicas⁵ extraídas del RGPD (Reglamento General de Protección de Datos) europeo. En particular, Como ya se explicó anteriormente, recopilamos las disposiciones normativas del GDPR empleando el D-KB (Robaldo et al., 2020), que representa en LegalRuleML las normas del GDPR. El La versión actual del D-KB6 incluye 966 fórmulas en formato cosificado. Lógica de entrada/salida: 271 obligaciones, 76 permisos y 619 Reglas constitutivas. Como se explica en Robaldo et al. (2020), el El número de reglas constitutivas es mucho mayor que el de permisos. y obligaciones porque se necesitan reglas constitutivas para desencadenar inferencias especiales para las reglas modeladas. En otras palabras,

Article 5

Principles relating to processing of personal data

1. Personal data shall be:

(a) processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');

(b) collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation');

(c) adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation');

[...]

Como puede verse en el texto anterior, el párrafo 1 del artículo 5 es una lista compuesta por una parte introductoria ("Los datos personales deberán ser:") y diferentes puntos. Para simplificar, sólo reportamos el Los tres primeros puntos del apartado 1, a saber, la letra a, la letra b y la letra

⁵ Para la selección de las disposiciones, excluimos el preámbulo y conclusión del documento legal principal del RGPD, manteniendo así solo las disposiciones dentro del cuerpo del RGPD. Estas disposiciones son generalmente párrafos o puntos de lista y, en ocasiones, pueden Consta de más de una frase.

⁶ El D-KB se puede descargar gratuitamente desde: https://github.com/dapreco/daprecokb/blob/master/gdpr/rioKB_GDPR.xml.

⁷ La representación LegalDocML del RGPD puede estar actualmente accedido y descargado desde <https://github.com/guerret/lu.uni.dapreco.parser/blob/master/resources/akn-act-gdpr-full.xml>. <https://eur-lex.europa.eu/eli/reg/2016/679/oj#d1e1807-1-1>.

Tabla 2 – Resultados para los cuatro escenarios de clasificación. Métrica de evaluación: puntuación F1 (puntuación F1 ponderada para los escenarios multiclase 3 y 4). RI indicó la mejora relativa en puntos decimales en comparación con la mejor línea de base.

	Models from Liga and Palmirani 2022 [16]			Our models		RI
	BERT	DistilBERT	LegalBERT	GPT-3 (Prompt 1)	GPT-3 (Prompt 2)	
Scenario 1	.86	.88	.82	.90	.91	+3
Scenario 2	.88	.92	.88	.90	.93	+1
Scenario 3	.88	.84	.85	.94	.93	+6
Scenario 4	.78	.80	.81	.91	.93	+12

C. Desde el punto de vista del lenguaje natural, cada oración deóntica se divide entre la parte introductoria (que contiene el verbo deóntico principal "shall") y el texto de cada punto. Si bien la parte introductoria contiene el verbo deóntico principal, la información deóntica real está contenida en cada punto.

La formalización de LegalDocML para el punto a es la siguiente:

```
<article eId="art_5">
  <num> Article 5 </num>
  <heading eId="art_5_heading">
    Principles relating to processing of personal data
  </heading>
  <paragraph eId="art_5_para_1">
    <num> 1. </num>
    <intro> <p> Personal data shall be: </p> </intro>
    <point eId="art_5_para_1_content_list_1_point_a">
      <num> (a) </num>
      <content>
        <p> processed lawfully, fairly and in a transparent manner in relation
        to the data subject ('lawfulness, fairness and transparency');
      </p>
      </content>
    </point>
  </...>
```

En el D-KB, que emplea LegalRuleML, se puede encontrar una serie de elementos <Legal Reference>, que contienen la parte estructural donde se ubican las fórmulas deónticas, referenciadas mediante el uso de la convención de nomenclatura LegalDocML9. Por ejemplo, la referencia del punto a mencionado anteriormente está codificada en el D-KB de la siguiente manera:

```
<LegalReference
  refersTo = "gdprC2A5P1p1ref">
```

```
refID = "GDPR:
art-5-para-1-content-list-1-point-a"/>
```

en el que el atributo "refersTo" indica el ID interno de

la referencia, y el atributo "refID" indica el ID externo de la referencia utilizando la convención de nomenclatura LegalDocML. El prefijo "GDPR" significa el uri LegalDocML del GDPR, es decir, "/akn/eu/act/regulation/2018-05-25/eng@2018-05-25/main#".

A su vez, este elemento <LegalReference> se asocia a su grupo objetivo de enunciados lógicos, que recoge el grupo de fórmulas lógicas relacionadas con esta referencia legal (en este caso, relacionadas con el punto a del primer párrafo del artículo 5). . Dicha asociación se modela de la siguiente manera:

```
<Association>
  <appliesSource keyref="#gdprC2A5P1p1ref" />
  <toTarget keyref="#statements1" />
</Association>
```

Donde el atributo "keyref" del destino conecta la fuente con la colección de declaraciones cuyo atributo "clave" es "declaraciones1":

```
<Statements key="statements1">

  <ConstitutiveStatement key="statements1Formula1">
    <Rule closure="universal">
      <if>[...]</if>
      <then>[...]</then>
    </Rule>
  </ConstitutiveStatement>

  <ConstitutiveStatement key="statements1Formula2">
    <Rule closure="universal">
      <if>[...]</if>
      <then>[...]</then>
    </Rule>
  </ConstitutiveStatement>

</Statements>
```

Es importante subrayar que cada enunciado en lenguaje natural puede tener múltiples fórmulas en el ámbito lógico.

⁹ <https://docs.oasis-open.org/legaldocml/akn-nc/v1.0/csprd01/akn-nc-v1.0-csprd01.html>.

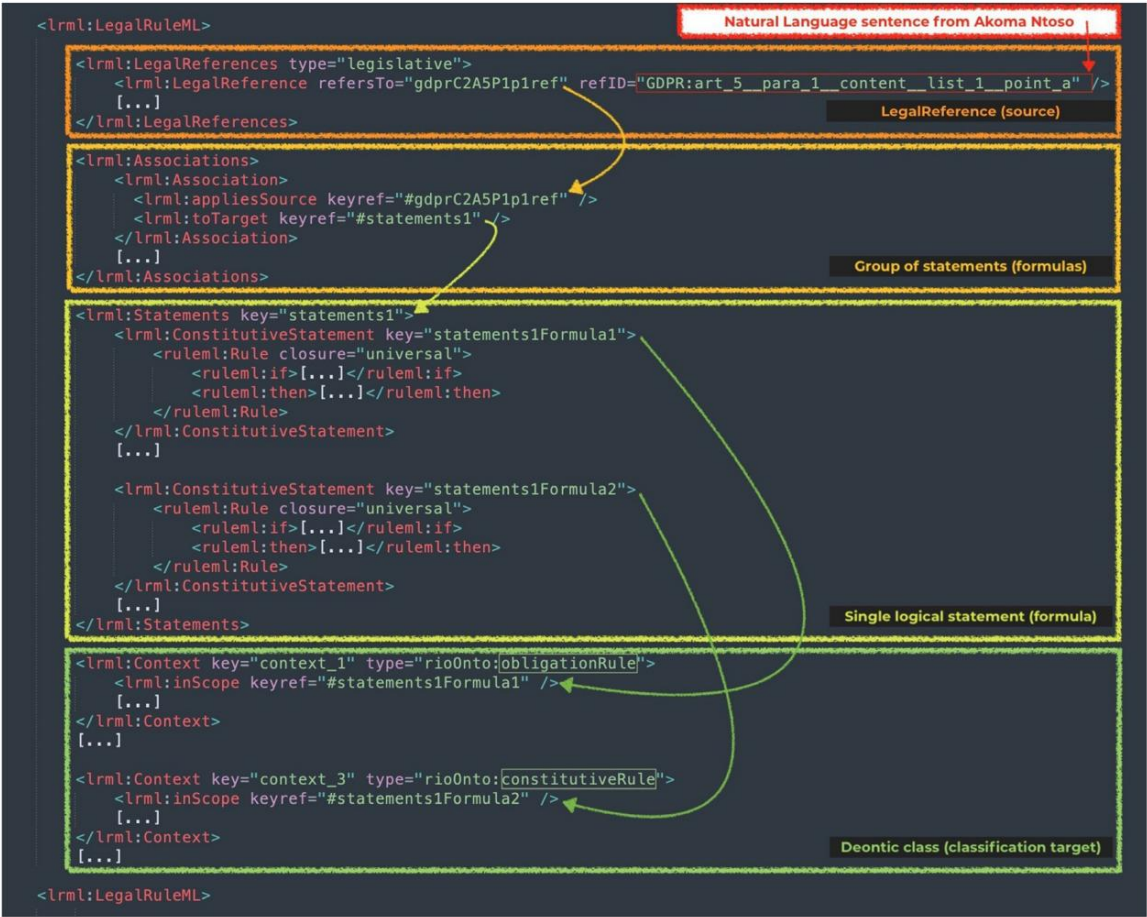


Fig. 5 – Descripción del proceso de extracción de clases de LegalDocML y DAPRECO.

Por esta razón, el elemento `<Statements>` aquí muestra una colección de dos fórmulas lógicas.

Para finalmente asociar la porción de lenguaje natural extraída de LegalDocML a una clase relacionada con la lógica esfera (es decir, la clase deóntica), uno debe mirar el `<Contexto>` elementos que están relacionados con las dos fórmulas que encontramos.

```
<Context key="context_1"
type="rioOnto:obligationRule">
  <inScope
keyref="#statements1Formula1" />
  [...]
</Context>

<Context key="context_3"
type="rioOnto:constitutiveRule">
  <inScope
keyref="#statements1Formula2" />
  [...]
</Context>
```

Como puede verse en el texto anterior, la primera fórmula (que se llama aquí “declaraciones1Fórmula1”) está asociado con el clase ontológica “obligationRule”, mientras que la segunda fórmula (que se llama “declaraciones1Fórmula2”) está asociada con la clase ontológica “Regla constitutiva”. Esto significa que el pieza de lenguaje natural expresada en el inciso a del primer párrafo del art. 5 del RGPD contiene, a nivel lógico, una regla constitutiva y una regla de obligación.

La Fig. 6 muestra la serie completa de pasos desde la esfera del lenguaje natural (ubicada en LegalDocML) hasta la esfera lógica. (es decir, la formalización LegalRuleML) donde las clases deónticas Están localizados. La figura explica paso a paso cómo la combinación de LegalDocML y LegalRuleML nos ayudó en la extracción de datos etiquetados y anotados.

6. Configuración y resultados del experimento

En lo que respecta a los entornos experimentales, el conjunto de datos se dividió en conjuntos de entrenamiento y validación, con una puntuación de 80/20 dividir10. Además, utilizamos Ada como motor para GPT-3. empleó Ada, la opción estándar para tareas de clasificación.

¹⁰ Tenga en cuenta que esta proporción fue ligeramente diferente en (Liga y Palmi-rani, 2022).

Tabla 3 – Resultados para los cuatro escenarios de clasificación. Métrica de evaluación: Precisión. RI indicó la mejora relativa en puntos decimales en comparación con la mejor línea de base.

	Models from Liga and Palmirani 2022 [16]			Our models		RI
	BERT	DistilBERT	LegalBERT	GPT-3 (Prompt 1)	GPT-3 (Prompt 2)	
Scenario 1	.86	.88	.82	.94	.94	+6
Scenario 2	.88	.92	.88	.95	.96	+4
Scenario 3	.87	.83	.84	.94	.93	+7
Scenario 4	.75	.78	.76	.91	.93	+15

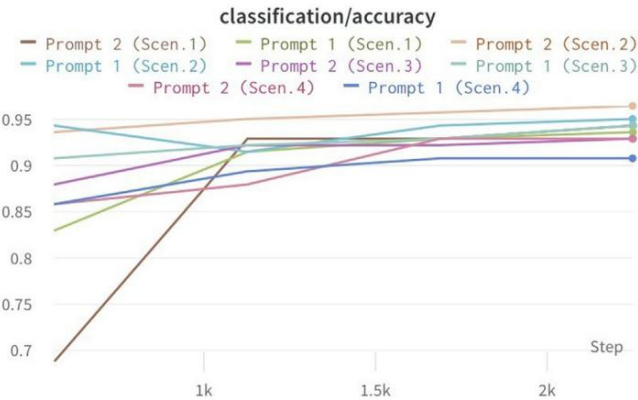


Fig. 6 – Precisión para los cuatro escenarios de clasificación considerando las dos indicaciones.

Además, notamos empíricamente que Ada de GPT-3 supera a Davinci en tareas de clasificación simples, mientras que Davinci es más apropiado en tareas generativas. Informamos los resultados de BERT (Devlin et al.), distilBERT (Sanh et al.) y LegalBERT.

(Chalkidis et al.), tomado de Liga y Palmirani (2022a), que utilizamos como referencia, y también informamos nuestros dos modelos GPT-3 perfeccionados. Estos dos modelos de lenguaje afinados se lograron después de 4 épocas de ajuste. En lo que respecta a los hiperparámetros, establecemos el multiplicador de la tasa de aprendizaje en 0,1, el peso de pérdida rápida en 0,01 y el tamaño del lote en 1. Los resultados finales del conjunto de validación se informan en las Tablas 2 y 3, donde se puede observar que todos los modelos GPT-3 ajustados superan los resultados anteriores tomados de Liga y Palmirani (2022a), tanto en términos de puntuaciones de F1 (Tabla 2) como en términos de precisión (Tabla 3). Higos. 6 y 7 muestran precisión y puntuaciones F1 para nuestros modelos ajustados.

Respecto al escenario 1 (es decir, la clasificación binaria entre reglas y no reglas) logramos .91 de puntuación F1 con la pregunta 2 y .90 con la pregunta 1, mientras que el mejor resultado anterior fue .88 con DistilBERT.

En el segundo escenario, que consiste en realizar una clasificación binaria entre reglas deónticas (obligaciones y permisos) vs reglas no deónticas (es decir, todo lo que no es obligación ni permiso), se alcanzó GPT-3. 93 de puntuación de F1, contra .92 (DistilBERT). Este es el único escenario en el que la línea base compitió con GPT-3.

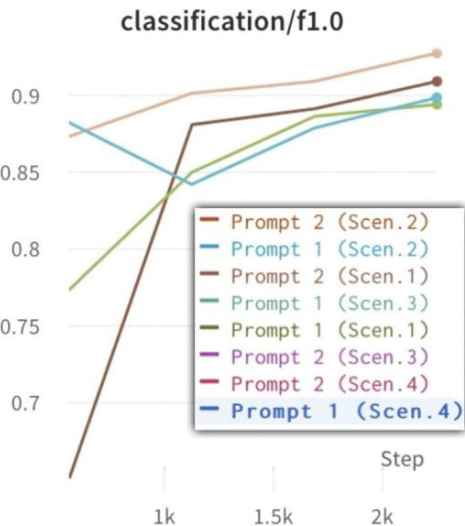
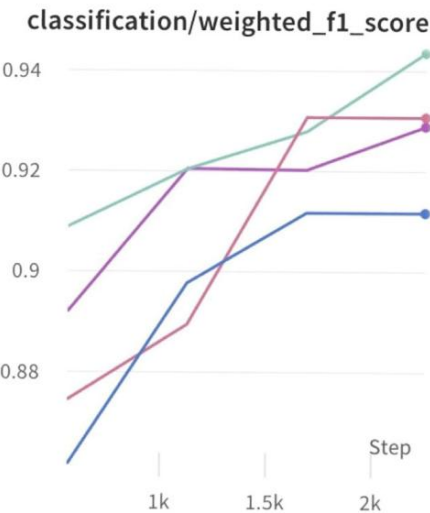


Fig. 7 – Puntajes F1 para escenarios multiclase (escenarios 3 y 4) y para clasificaciones binarias (escenarios 1 y 2) a la izquierda y a la derecha respectivamente.

Respecto al tercer escenario, consistente en un multiclase clasificación que implica obligaciones y permisos contra “no deóntico” (es decir, todo lo demás que no es una obligación ni un permiso), GPT-3 alcanzó .94 de puntuación F1, contra .88 (BERT).

La mejora más significativa está relacionada con la escenario más difícil, el cuarto, que implica la clasificación más granular, considerando todos los disponibles clases (obligación, permiso, regla constitutiva, no regla). Mientras que LegalBERT, el resultado más alto anterior, logró .81 de Puntuación F1 ponderada, nuestros modelos GPT-3 ajustados alcanzaron 0,93 de Puntuación F1 ponderada, más de un decimal.

Como puede verse en la [Tabla 3](#), la mejora relativa de La línea de base es más evidente cuando se considera la precisión como métrica de evaluación.

7. Conclusión y trabajos futuros.

En este trabajo, demostramos que los LLM generativos autorregresivos como GPT-3 puede superar a los LLM con codificación automática como BERT logrando mayores desempeños en la tarea de Clasificación de Normas Jurídicas (LRC). Realizamos LRC en cuatro escenarios de clasificación multiclase, que involucran obligaciones, permisos, reglas constitutivas y “no reglas” (es decir, disposiciones legales que no contener algún tipo de regla).

Es importante destacar que nuestro trabajo muestra cómo se acerca la IA híbrida. puede combinar con éxito Inteligencia Artificial simbólica y subsimbólica. Por un lado, proporcionamos expertos legales. conocimiento codificado en formatos LegalXML como LegalRuleML y Akoma Ntoso. Por otro lado, aprovechamos el poder de GPT-3, posiblemente el modelo de lenguaje más potente actualmente disponibles para afinar en el campo de la Inteligencia Artificial. Este enfoque de IA híbrida muestra el potencial de combinar enfoques de arriba hacia abajo (basados en el conocimiento) con enfoques de abajo hacia arriba. métodos (basados en datos).

Hasta donde sabemos, este documento es el primero en ajustar GPT-3 (actualmente uno de los modelos más poderosos y discutidos intensivamente) en datos legales. Aunque nuestro trabajo muestra las impresionantes capacidades de los LLM para aprender de pequeñas cantidades de datos, así como la superioridad de GPT-3 En comparación con otros LLM, se requieren más experimentos para validar. estas tendencias utilizando diferentes conjuntos de datos legales.

En trabajos futuros, planeamos probar GPT-3 con un rango más amplio. de indicaciones para evaluar su desempeño en tareas más complejas y detalladas relacionadas con LRC y la clasificación de modalidad deóntica. El potencial para interrogar mejor los detalles de estos datos puede ser particularmente intrigante cuando se combinan con otras tareas de PNL como SRL ([Humphreys et al., 2021](#)). Otra perspectiva interesante es la creación de sistemas expertos que puedan identifique los requisitos legales automáticamente y verifique su cumplimiento en consecuencia.

Si bien el último GPT-4 no está disponible actualmente para realizar ajustes (OpenAI anunció su próxima accesibilidad), planear realizar una serie de experimentos de aprendizaje de pocas tomas con GPT-4 y compare los resultados con el escenario de ajuste actual. También planeamos emplear otros LLM como como MPT ([Team, 2023](#)), Dolly-v2 ([Conover et al., 2023](#)), BLOOM ([Scao et al.](#)), LLaMA ([Touvron et al.](#)).

Declaración de intereses en competencia

Los autores declaran que no tienen intereses financieros en conflicto ni relaciones personales conocidas que puedan haber influido en el trabajo presentado en este artículo.

Disponibilidad de datos

Los datos estarán disponibles previa solicitud.

Reconocimiento

Davide Liga contó con el apoyo del proyecto INDIGO, que está apoyado financieramente por el Programa Conjunto de Investigación NORFACE sobre Gobernanza Democrática en una Era Turbulenta y cofinanciado por AEI, AKA, DFG y FNR y la Comisión Europea a través de Horizonte 2020 en virtud del acuerdo de subvención no [822166](#).

Livio Robaldo ha contado con el apoyo de la operación Legal Innovation Lab Wales dentro del programa Hillary de la Universidad de Swansea. Facultad de Derecho Rodham Clinton. La operación ha sido financiada parcialmente por el Fondo Europeo de Desarrollo Regional a través de el Gobierno de Gales.

referencias

Ashley K.D. Inteligencia artificial y analítica jurídica: nuevas herramientas para la práctica del derecho en la era digital. *Prensa de la Universidad de Cambridge*; 2017.

Athan T, Boley H, Governatori G, Palmirani M, Paschke A, Wyner A. Oasis LegalRuleML. *Actas del decimocuarto conferencia internacional sobre inteligencia artificial y derecho*; 2013. pág. 3–12.

Boella G, Caro LD, Rispoli D, Robaldo L. Un sistema de clasificación texto de múltiples etiquetas en EuroVoc. *ICAIL. ACM*; 2013. pág. 239–40.

Boella G, Di Caro L, Humphreys L, Robaldo L, Rossi R, van der Torre L. Eunomos, un documento jurídico y un conocimiento sistema de gestión para la web para proporcionar información relevante y confiable e información actualizada sobre la ley. *Ley Artif Intell* 2016;4.

Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S, Herbert-Voss A, Krueger G, Henighan T, Child R, Ramesh A, Ziegler D, Wu J, Winter C, Hesse C, Chen M, Sigler E, Litwin M, Gray S, Ajedrez B, Clark J, Berner C, McCandlish S, Radford A, Sutskever I, Amodei D. Los modelos de lenguaje son de pocas tomas. estudiantes. En: Larochelle H, Ranzato M, Hadsell R, Balcan M, Lin H, editores. *Avances en el procesamiento de información neuronal sistemas*. Curran Asociados, Inc.; 2020. pág. 1877–901. vol. 33URL. <https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfc4967418bfb8ac142f64a-Paper.pdf>.

Chalkidis I., Androutsopoulos I., Michos A. Obligación y extracción de prohibición utilizando RNNs jerárquico, preimpresión de arXiv arXiv:1805.03871.

Chalkidis I., Fergadiotis M., Malakasiotis P., Aletras N., Androutsopoulos I. Legal-bert: los muppets sacados de la ley escuela, preimpresión de arXiv arXiv:2010.02559.

Conover M., Hayes M., Mathur A., Xie J., Wan J., Shah S., Ghodsi A., Wendell P., Zaharia M., Xin R. Free Dolly: presentando el La primera película de cine verdaderamente abierta y adaptada a la instrucción del mundo (2023). URL <https://www.databricks.com/blog/2023/04/12/dolly-primer-a-instrucción-abierta-comercialmente-viable-sintonizada-llm>

- de Maat E, Winkels R. Clasificación automática de sentencias en las leyes holandesas. *Conocimientos jurídicos y sistemas de información*. Prensa IOS; 2008. pág. 207–16.
- Devlin J., Chang M.-W., Lee K., Toutanova K. Bert: entrenamiento previo de transformadores bidireccionales profundos para la comprensión del lenguaje, preimpresión de arXiv arXiv:1810.04805.
- Dragoni M, Villata S, Rizzi W, Governatori G. Combinando naturales Enfoques de procesamiento del lenguaje para la extracción de reglas de documentos legales. En: Pagallo U, Palmirani M, Casanovas P, Sartor G, Villata S, editores. *La IA se acerca a la complejidad de los sistemas jurídicos*. Cham: Springer International Publishing; 2018. pág. 287–300.
- Gao X, diputado Singh. Extraer relaciones normativas de contratos comerciales. *Actas de la conferencia internacional de 2014 sobre agentes autónomos y sistemas multiagente*; 2014. pág. 101–8.
- Gómez-Pérez JM, Denaux R, García-Silva A. Procesamiento híbrido del lenguaje natural: una introducción. Cham: Springer International Publishing; 2020. pág. 3–6.
- Humphreys L, Boella G, van der Torre L, Robaldo L, Di Caro L, Ghanavati S, Muthuri R. Población de ontologías jurídicas mediante etiquetado de roles semánticos. *Ley Artif Intell* 2021;29:171–211.
- Joshi V, Anish PR, Ghaisas S. Adaptación de dominio para un Clasificación automatizada de modalidades deónticas en contratos de ingeniería de software. *Actas de la 29.ª reunión conjunta de la ACM sobre la conferencia y simposio europeo de ingeniería de software sobre los fundamentos de la ingeniería de software*; 2021. pág. 1275–80.
- Kiyavitskaya N, Zeni N, Breaux TD, Antoní AI, Cordy JR, Mich L, Mylopoulos J. Automatización de la extracción de derechos y obligaciones para el cumplimiento normativo. *Congreso internacional sobre modelado conceptual*. Saltador; 2008. pág. 154–68.
- Liga D, Palmirani M. Transferir aprendizaje para el gobierno deóntico Clasificación: el caso de estudio del RGPD. *Congreso internacional sobre conocimientos jurídicos y sistemas de información*. Silla fácil; 2022a. pág. 2022 Saarbrücken 14-16 de diciembre.
- Liga D, Palmirani M. Clasificación de oraciones deónticas utilizando clasificadores de núcleo de árbol. *Sistemas y aplicaciones inteligentes: actas de la conferencia de sistemas inteligentes de 2022 (IntelliSys)*. Springer International Publishing Cham; 2022b. pág. 54–73 Volumen 1.
- Liga D, Palmirani M. Transferir el aprendizaje con incrustaciones de oraciones para la clasificación de evidencia argumentativa. *Actas del vigésimo taller sobre modelos computacionales de argumento natural*, 2020.
- Liga D. Inteligencia artificial híbrida para extraer patrones y reglas a partir de textos argumentativos y jurídicos.
- Neill JO, Buitelaar P, Robin C, Brien LO. Clasificación oracional Modalidad en lenguaje jurídico: un caso de uso en regulaciones, leyes y directivas financieras. *Actas de la 16ª edición de la conferencia internacional sobre inteligencia artificial y derecho*; 2017. pág. 159–68.
- Nguyen M, Nguyen T, Tran V, Nguyen H, Nguyen L, Satoh K. Aprender a asignar el RGPD a la representación lógica en DAPRECO-KB. *ÁCIDOS*. Saltador; 2022. pág. 442–54 (1), vol. 13757 de *Apuntes de conferencias en Ciencias de la Computación*.
- Palmirani M, Vitali F. Akoma-Ntoso para documentos legales. *XML legislativo para la web semántica*. Saltador; 2011. pág. 75-100.
- Radford A., Narasimhan K. Mejora de la comprensión del lenguaje mediante formación previa generativa, 2018.
- Robaldo L, Villata S, Wyner A, Grabmair M. Introducción para Inteligencia artificial y derecho: número especial "Procesamiento del lenguaje natural para textos legales". *Ley Artif Intell* 2019;27(2):113–15.
- Robaldo L, Bartolini C, Palmirani M, Rossi A, Martoni M, Lenzini G. Formalizar las disposiciones del RGPD en una lógica de E/S reificada: la base de conocimientos de DAPRECO. *Informe J Logic Lang* 2020;29(4):401–49.
- Robaldo L, Bartolini C, Lenzini G. La base de conocimientos dapreco: representando el RGPD en LegalRuleML. *Actas de la duodécima conferencia de evaluación y recursos lingüísticos*; 2020. pág. 5688–97.
- Robaldo L. Hacia la verificación de cumplimiento en lógica de E/S cosificada vía SHACL. En: Maranhão J, Wyner AZ, editores. *Proc. de la 18ª conferencia internacional sobre inteligencia artificial y derecho (ICAIL)*. ACM, 2021.
- Rodríguez-Doncel V, Palmirani M, Araszkievicz M, Casanovas P, Pagallo U, Sartor G. Introducción: Un marco regulatorio híbrido y una arquitectura técnica para una IA explicable y centrada en el ser humano, en: *AI Approaches to the Complexity of Legal Systems XI-XII*. Saltador; 2020. pág. 1–11.
- Sanh V., Debut L., Chaumond J., Wolf T. Distilbert, una versión destilada de bert: más pequeña, más rápida, más barata y más ligera, preimpresión de arXiv arXiv:1910.01108.
- Scao TL, Fan A., Akiki C., Pavlick E., Ilic S., Hesslow D., Castagne R., Luccioni AS, Yvon F., Galle M., et al. Bloom: un modelo de lenguaje multilingüe de acceso abierto con parámetros 176b, preimpresión de arXiv arXiv:2211.05100.
- Shaghaghian S, Feng LY, Jafarpour B, Pogrebnikov N. Personalización de modelos de lenguaje contextualizados para revisiones de documentos legales. *2020 Conferencia internacional IEEE sobre big data (Big Data)*. IEEE; 2020. pág. 2139–48.
- Song D, Vold A, Madan K, Schilder F. Clasificación de documentos legales con múltiples etiquetas: un enfoque basado en el aprendizaje profundo con atención a las etiquetas y capacitación previa en un dominio específico. *Sistema Inf* 2022;106.
- Sun X, Robaldo L. Sobre la complejidad de la lógica de entrada/salida. *J Appl Lógica* 2017;25:69–88.
- Team MN Presentamos mpt-7b: un nuevo estándar para películas de código abierto utilizables comercialmente, consultado: 2023-05-05 (2023). URL www.mosaicml.com/blog/mpt-7b
- Touvron H., Lavril T., Izacard G., Martinet X., Lachaux M.-A., Lacroix T., Rozière B., Goyal N., Hambro E., Azhar F., et al. Llama: modelos de lenguaje básico abiertos y eficientes, preimpresión de arXiv arXiv:2302.13971.
- Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser L., Polosukhin I. La atención es todo lo que necesita, 2017. URL <https://arxiv.org/pdf/1706.03762.pdf> Wälti B, Muhr J, Glaser I, Bonczek G, Scepankova E, Matthes F. Clasificación de normas jurídicas con aprendizaje automático activo. *JÚRIX*; 2017. pág. 11–20.
- Wyner A, Peters W. Sobre la extracción de reglas de las regulaciones. *Conocimientos jurídicos y sistemas de información*. Prensa IOS; 2011. pág. 113-22.