Inteligencia artificial aplicada al estudio del derecho: análisis computacional de la jurisprudencia de casación penal

David Mielnik[1]

Edgar Altszyler[2]

Introducción

Los avances tecnológicos siempre han despertado el interés de las y los juristas. Incluso los primeros registros que tenemos de códigos de leyes y discusiones doctrinarias —el Código de Hammurabi (1754 a. e. c.) o el Talmud (200 e. c.), por ejemplo— contienen numerosas referencias a desarrollos y descubrimientos de sus respectivas épocas, tales como técnicas agriculturales, ejercicios de la medicina, uso de pesos y medidas, armas y herramientas cada vez más sofisticadas[3]. La inteligencia artificial (IA), uno de los desarrollos más importantes de los últimos años y motor de lo que algunos ya llaman la 4ta Revolución Industrial[4], no es la excepción.

A lo largo de los siglos, esa relación entre derecho y tecnología ha tenido, en general, una dirección única: la de *la tecnología mirada a través del derecho*. Preocupados tanto por maximizar la utilidad social de la ciencia, como por minimizar sus efectos perjudiciales, nuestra inclinación habitual como abogados y abogadas es discutir las formas más eficientes y justas de regulación de sus avances. Ese impulso es desde luego sano y deseable, porque tecnologías como la IA ciertamente poseen un potencial disruptivo enorme, y están llamadas a alterar radicalmente casi todos los aspectos de la vida en sociedad. Muchos de los capítulos de la obra colectiva que implica este Tratado representan lo mejor de esas reflexiones acerca de la regulación.

Con esta contribución, sin embargo, queremos dar vuelta a esa relación y mostrar que los avances producidos en inteligencia artificial permiten que hoy en día podamos también generar *una mirada del derecho a través de la tecnología*. Después de todo, el derecho mismo se está convirtiendo —cada vez más rápidamente— en un fenómeno frente al que nuestras herramientas tradicionales de análisis legal comienzan a mostrar

[1] Escuela de Derecho, Universidad Torcuato Di Tella (david.m@utdt.edu).

[2] Laboratorio de Inteligencia Artificial Aplicada, Universidad de Buenos Aires (edgaralts@gmail.com).

[3] Por ejemplo, leyes 53, 55 y 108 del Código de Hammurabi; Tratados de Bejorot (37-38), Gitin (67b) y Shabat (63a) del Talmud Babli.

[4] VELARDE, G., "Artificial Intelligence and Its Impact on the Fourth Industrial Revolution: A Review", ArXiv, 2020.

limitaciones: un fenómeno de *big data*[5] que se expresa en cientos de miles de sentencias judiciales, decisiones administrativas, textos normativos y opiniones doctrinarias. Así, al darnos las herramientas para dar cuenta de grandes volúmenes de información, tecnologías emergentes como la IA comienzan a ser imprescindibles para poder observar muchos de los patrones, regularidades y tendencias que hacen a las reglas y principios jurídicos, y que de otra forma permanecerían ocultos —latentes—, diseminados a lo largo de vastas colecciones de textos que el intelecto humano, por sí mismo, es incapaz de procesar.

En particular, en las páginas que siguen ilustraremos esta idea aplicando al derecho argentino un algoritmo de procesamiento del lenguaje mediante inteligencia artificial. Mostraremos cómo, a través de esta clase de sistemas, podemos desarrollar una mirada jurídica única y novedosa, formada a partir del procesamiento de, literalmente, miles de decisiones judiciales.

Nuestra presentación está estructurada de la siguiente manera. En la primera sección (i.) presentaremos los contornos del *Análisis Computacional del Derecho*, el área de estudios interdisciplinarios que promueve la adopción de tecnologías digitales para amplificar nuestras capacidades de análisis legal. Aquí precisaremos también qué es exactamente la inteligencia artificial, en el sentido en el que utilizamos hoy en día el término, con toda su carga disruptiva. Concluiremos esta sección relevando algunos trabajos recientes que han empleado IA para estudiar derecho y ciencias sociales —lo que hasta ahora ha ocurrido, en general, sólo en Estados Unidos y Europa—.

Dedicaremos la segunda sección (ii.) a explicar con mayor detalle algunas intuiciones fundamentales sobre el funcionamiento del algoritmo de modelado de tópicos (o *topic modelling*) mediante IA que utilizaremos para nuestro análisis de la jurisprudencia argentina. Dedicaremos la tercera sección (iii.) al principal aporte que queremos realizar con esta presentación: la demostración del uso de la IA para el estudio del derecho mediante el procesamiento algorítmico de más de 80.000 sentencias judiciales, pronunciadas por la Cámara Federal de Casación Penal desde 1992 hasta 2022. La ocasión se presenta especialmente oportuna ya que recientemente, el 22 de diciembre de 2022, aquel tribunal —el de mayor jerarquía del ordenamiento federal con competencia exclusivamente penal— cumplió 30 años de funcionamiento ininterrumpido. Nuestro estudio, así, pretende mostrar cómo la IA nos puede ayudar a sintetizar, en pocas páginas, una porción importante de la historia jurisdiccional del país. Dedicaremos la última sección (iv.) a exponer nuestras conclusiones.

---

[5] ALSCHNER, W., "The Computational Analysis of International Law", en DEPLANO, R. y TSAGOURIAS, N. (eds), *Research Methods in International Law: A Handbook*, Edward Elgar Publishing, 2019.

1. El Análisis Computacional del Derecho

1.1 Cómo abarcar un derecho inabarcable

La importancia de mantener un conocimiento actualizado de los principales temas que se ventilan en los tribunales no puede minimizarse. Este conocimiento no sólo ayuda a que litigantes y jueces puedan navegar el cada vez más complejo paisaje legal, elaborar estrategias sólidas e informar sus propios procesos decisionales, sino que también es una herramienta imprescindible, por ejemplo, para que los órganos legislativos y de gobierno puedan entender el impacto social que tienen las leyes y decisiones administrativas en la práctica, y para que las agencias de seguridad tengan un conocimiento preciso de los límites de su autoridad.

El ecosistema legal, administrativo y jurisprudencial, empero, se ha vuelto endemoniadamente complejo en los últimos años, y comienza a presentar las dificultades de acceso a la información propias de los fenómenos de *big data*[6], frente a los que nuestras herramientas tradicionales de análisis tienen evidentes limitaciones. Los tribunales, por caso, publican literalmente decenas de miles de decisiones por mes, que a su vez se suman a un verdadero laberinto de regulaciones y disposiciones de toda jerarquía normativa. ¿Cómo dar cuenta de un derecho tan vasto?

En los últimos cinco o diez años, la democratización del acceso a recursos informáticos cada vez más poderosos ha impulsado a investigadoras e investigadores del derecho a intentar superar estos desafíos mediante la adopción de nuevas tecnologías, que pueden ser utilizadas para descubrir conocimiento mediante la explotación de grandes volúmenes de información, de formas que hasta hace poco parecían imposibles. Este abordaje interdisciplinario recibe habitualmente el nombre de Análisis Computacional del Derecho (ACD)[7]: una forma de ciencia de datos aplicada, que combina tecnología de la información con saberes jurídicos para generar nuevo conocimiento a partir de la identificación de patrones y regularidades que están diseminados a lo largo de decenas de miles de textos jurídicos y que, por esa razón, nos resultan inaccesibles a simple vista[8].

El estudio computational que presentamos en la Sección 3, realizado mediante el procesamiento de más de 100.000 sentencias judiciales de Argentina, se inscribe en esta corriente de investigación, que es a la vez tributaria de los llamados Estudios

---

[6] El Observatorio Nacional de Big Data lo define como "una cantidad de información que sólo se puede procesar mediante el uso de herramientas digitales, y que sirve para responder preguntas a través del análisis de enormes volúmenes de datos". Ver https://www.argentina.gob.ar/grupo-de-trabajo/observatorio-nacional-big-data/observatorio-big-data/que-es-big-data.

[7] MIELNIK, D., "Análisis Computacional del Derecho Argentino", Revista Argentina de Teoría Jurídica, vol. 23, 2022; ALSCHNER, W., "The Computational Analysis of International Law"; y VARSAVA, N., "Computational Legal Studies, Digital Humanities, and Textual Analysis", *Computational Legal Studies*, Edward Elgar Publishing, 2020.

[8] En este sentido, ver MIELNIK, D., "Ciencia de Datos y Análisis Del Derecho Penal", Revista de Derecho Penal y Criminología, vol. 12, 2022.

Legales Empíricos[9] y de las Humanidades Digitales —la disciplina que estudia la convergencia entre las ciencias humanísticas y la tecnología[10]—. En efecto, a pesar de su relativa novedad, lo cierto es que con el Análisis Computacional del Derecho la profesión legal comienza a dar sus primeros pasos en un camino que numerosas disciplinas, incluyendo ciencias sociales como la historia o la politología, en realidad ya vienen recorriendo hace tiempo[11].

1.2 Algunas precisiones sobre inteligencia artificial

Entre las premisas que fundamentan el ACD se encuentra la observación de que el derecho es, entre otras cosas, un conjunto de datos susceptibles de ser analizados y procesados por una máquina[12]. Así, la IA es una de las tecnologías digitales más interesantes para las y los analistas computacionales del derecho. Quizás por la gran atención que atrae el término, sin embargo, muchas veces es utilizado de manera ambigua, casi como eslogan publicitario, opacando y dificultando el debate respecto de su uso y regulación. Para entender correctamente cuál es el aporte que puede hacer esa tecnología al estudio y la comprensión del derecho, conviene entonces hacer algunas precisiones.

En un sentido muy amplio, en efecto, se habla de Inteligencia Artificial para describir cualquier proceso que una computadora pueda realizar de forma automática. Esa fue en efecto la definición con la que se la conoció, en los años '50 y '60[13]. Durante esos años, su desarrollo estuvo enfocado en lograr traducir la capacidad humana para tomar decisiones en reglas estáticas, pre-escritas por las y los analistas, que necesariamente debían prever todas las posibles demandas a las que el sistema habría de enfrentarse. Es que, por muy sofisticados que sean, tales sistemas carecen de la habilidad para aprender de los datos externos de manera autónoma[14] y están limitados al cumplimiento rígido y mecánico de un programa (que funciona esencialmente como un guión o una receta de cocina).

Más allá de un modesto éxito inicial en la producción de sistemas capaces de responder ciertas preguntas técnicas (siempre, claro, que esas preguntas hubieran sido expresamente previstas y contestadas ya por los creadores del artefacto al programarlo), o incluso jugar algunos juegos simples, cuyas reglas pudieran ser

---

[9] Ver, por ejemplo, EISENBERG, T., "The Origins, Nature, and Promise of Empirical Legal Studies and a Response to Concerns", SSRN Journal, 2010.

[10] BURDICK, A., DRUCKER, J., LUNENFELD, P., PRESNER, T., y SCHNAPP, J., "Digital_Humanities", MIT Press, 2016.

[11] Ver, por ejemplo, GRIMMER, J. y STEWART, B.M., "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts", Political Analysis, vol. 21, 2013.

[12] MIELNIK, D., "Análisis Computacional del Derecho Argentino", p. 6.

[13] LIGHTHILL, J., "Artificial Intelligence: A General Survey", 1973, p. 8.

[14] KAPLAN, A. y HAENLEIN, M., "Siri, Siri, in My Hand: Who's the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence", Business Horizons, vol. 62, 2019, p. 4.

formalizadas con relativa facilidad, ya a comienzo de los años '70, la investigación comisionada por el Consejo de Investigación Científica del Reino Unido al Prof. Sir James Lighthill[15] demostró, de manera lapidaria, que las limitaciones inherentes a ese paradigma harían altamente improbable que aquellas demostraciones de laboratorio pudieran algún día generalizarse en herramientas genuinamente útiles para la sociedad. En efecto, más allá de ciertas tareas puntuales y relativamente sencillas, el número de actividades que pueden reducirse a un conjunto finito de reglas formales parece ser más bien bajo.

El *Informe Lighthill,* de 1973, condujo al así llamado "invierno de la inteligencia artificial": un período que se extendió al menos hasta los años '90, caracterizado por una profunda desconfianza hacia la utilidad práctica de la IA y la virtual paralización del área de investigación.

El desarrollo que permitió escapar del invierno y reinventó la IA como una de las principales disciplinas detrás de la llamada 4ta Revolución Industrial fue la transición hacia un paradigma en el que la capacidad de *aprendizaje* por parte de un sistema es la condición *sine qua non* para que éste sea considerado genuinamente una "inteligencia artificial". Para ello, debe tener "la capacidad de interpretar datos externos correctamente, aprender de ellos y usar ese aprendizaje para realizar tareas y alcanzar objetivos específicos mediante procesos de adaptación flexibles"[16].

Así las cosas, sin desconocer sus orígenes ni el largo camino que nos trajo hasta aquí, para evitar confusiones conceptuales es importante tener en cuenta que, cuando hablamos hoy en día de la tecnología disruptiva que llamamos "inteligencia artificial", y de cómo parece estar cambiando el mundo a pasos agigantados, ciertamente no nos referimos a los sistemas rígidos de respuestas automatizadas que ya existían hace más de 70 años, sino, específicamente, a los así llamados algoritmos de *aprendizaje automático* o *machine learning*: sistemas inductivos que, en lugar de funcionar sobre la base de respuestas pre-configuradas e inamovibles, tienen la capacidad de ajustar automáticamente sus parámetros internos de funcionamiento y comportamiento, en respuesta a la exposición a información nueva[17].

Un ejemplo ayudará a entender la diferencia entre el paradigma de programación tradicional y el basado en el aprendizaje automático: supongamos que necesitamos un sistema que nos permita estimar cuánto nos costará un viaje en taxi. La forma tradicional de hacer eso es programar explícitamente una regla de conversión que tome en cuenta, por ejemplo, la bajada de bandera (hoy, $225 en la Ciudad de Buenos Aires), el precio por metro recorrido ($0,118) y la distancia a recorrer. Así, para estimar la tarifa de un viaje *V,* de 5.000 mts., calcularíamos $V = \$0{,}118 * 5.000 + \$225 = \$815$.

---

[15] LIGHTHILL, J., "Artificial Intelligence: A General Survey".

[16] KAPLAN, A. y HAENLEIN, M., "Siri, Siri, in My Hand", p. 3.

[17] DOMINGOS, P., "A Few Useful Things to Know about Machine Learning", Communications of the ACM, vol. 55, 2012, p. 81.

Bajo el paradigma del aprendizaje automático, por otro lado, lo que haríamos sería exponer a un algoritmo informático a un conjunto de viajes pasados, junto con sus correspondientes tarifas, y dejaríamos que el sistema *aprendiera* por sí mismo la regla de conversión implícita.

Pero entonces, ¿para qué molestarnos con el aprendizaje automático si la programación tradicional —explícita y basada en reglas— nos permite llegar al mismo resultado? La respuesta es que, más allá de ejemplos triviales como éste, en la mayoría de los casos relevantes *no* conocemos los patrones subyacentes a los datos y, por lo tanto, no nos es posible diseñar reglas pre-configuradas que los describan. Sin embargo, expuesto a un número suficiente de observaciones —en lo que llamamos la fase de *entrenamiento*— los algoritmos de *machine learning* pueden en general identificar propiedades tan diversas como las que permiten distinguir un perro de una medialuna, o los aspectos de un recurso de casación que mejor pueden anticipar su éxito procesal[18].

El resultado de entrenar un algoritmo de aprendizaje automático es la construcción de un *modelo*, esto es, "la representación matemática de las relaciones entre las variables que surgen de los datos a los que el algoritmo es expuesto"[19]. En otras palabras, una modelo es una definición operativa del objeto estudiado, formada a partir de la identificación de las regularidades que lo caracterizan.

Existen distintas clases de algoritmos que, a su vez, producen modelos con propósitos diferentes. Los más conocidos son los algoritmos de aprendizaje *supervisado*, que producen los llamados modelos *predictivos*. Reciben este nombre porque su entrenamiento se realiza con observaciones previamente anotadas, que actúan como guía para que el algoritmo pueda ir aprendiendo de sus propios errores, de un modo que recuerda en cierta medida al docente que corrige al estudiante equivocado[20]. Una vez que el algoritmo aprende a relacionar ciertas propiedades de los ejemplos de entrenamiento con una variable objetivo, se lo puede utilizar para identificar —*predecir*— el valor de esa variable en una observación no anotada.

La predicción de una cantidad continua —como el precio de mercado de un inmueble, a partir de sus características edilicias, o el tiempo que insumirá un proceso judicial, en función de los hechos del caso— es una tarea de *regresión*; mientras que la estimación de un valor discreto —identificar si una fruta es "pera" o "manzana", o si un recurso de apelación comparte las características de los recursos habitualmente "aceptados" o "rechazados"— es una tarea de *clasificación*[21].

A su vez, hay algoritmos de IA que realizan tareas de aprendizaje *no supervisado*, para las cuales no hay una etiqueta conocida *ex ante* que "supervise" y oriente al

---

[18] MIELNIK, D., "Análisis Computacional del Derecho Argentino", pp. 12–34.

[19] GRUS, J., "Data Science from Scratch", O'Reilly Media, 2015, pp. 141–142.

[20] MIELNIK, D., "Análisis Computacional del Derecho Argentino", pp. 9–13.

[21] JAMES, G.M., WITTEN, D., HASTIE, T., y TIBSHIRANI, R., "An Introduction to Statistical Learning", Springer, 2021, pp. 28–29.

algoritmo durante su entrenamiento. En su lugar, el aprendizaje *no supervisado* aprende las relaciones entre las propiedades mismas, y busca así desentrañar sus patrones subyacentes, latentes, normalmente con el propósito de agrupar las observaciones de algún modo coherente, con mayor o menor rigidez[22].

Una de las tareas habituales de los algoritmos de aprendizaje no supervisado aplicado al procesamiento del lenguaje consiste en la inferencia del *modelo de tópicos* que subyace a una muestra; esto es, la identificación de los temas presentes en una colección de documentos (habitualmente llamada *corpus*). Profundizaremos en el funcionamiento de estos algoritmos de IA no supervisada en la Sección 2. En efecto, en ellos se basa el estudio computacional sobre la jurisprudencia de casación penal que presentamos en este trabajo (Sección 3).

Antes de hacerlo, sin embargo, déjennos comentar algunos trabajos recientes que también han aprovechado la inteligencia artificial para realizar estudios en ciencias sociales y derecho.

Por cierto, al tiempo de presentar este capítulo, el mundo se encuentra todavía procesando el enorme potencial de productos como ChatGPT, basados en tecnologías de "inteligencia artificial generativa" —ampliamente discutida en este Tratado—.

Si bien esta nueva generación de IA está llamada a cambiar para siempre la manera en la que estudiamos y ejercemos el derecho, en estos momentos de "*hype*" es fácil sobreestimar los alcances de la tecnología, por lo que es más importante que nunca empezar a interesarnos por su funcionamiento interno y, sobre todo, conocer las limitaciones de este tipo de sistemas, que a primera vista pueden confundirnos y hacernos creer que son capaces de cumplir casi cualquier tarea[23]. Como cualquier tecnología, empero, la IA generativa es una herramienta más, y siempre es responsabilidad de las y los profesionales elegir la herramienta adecuada para la tarea correspondiente.

Por caso, GPT resulta especialmente apropiado para tareas que involucren la generación de texto —para eso fue creado— pero incluso su versión más moderna es probablemente muy ineficiente como herramienta para analizar el derecho "en perspectiva distante" mediante el descubrimiento de patrones y regularidades subyacentes en una colección de miles de extensos textos jurídicos —como nos proponemos hacer aquí—, especialmente cuando se trata de documentos con los que GPT no fue entrenado originalmente. Algunos estudios independientes también sugieren (para sorpresa de nadie) que las capacidades de GPT para cumplir

---

[22] JAMES, G.M. *et al.*, *An Introduction to Statistical Learning*, pp. 26–28.

[23] Para un ejemplo tragicómico del peligro de sobredimensionar las capacidades jurídicas de GPT, ver "Un abogado usó ChatGPT para buscar pruebas para defender a su cliente... pero resultaron ser inventadas", *La Nación* (30 de mayo de 2023, disponible en https://www.lanacion.com.ar/tecnologia/puede-fallar-un-abogado-uso-chatgpt-para-juntar-pruebas-para-defender-a-su-cliente-pero-resultaron-nid30052023/).

exitosamente consignas legales pueden haber sido exageradas en los folletos de promoción del producto[24].

1.3 Inteligencia artificial para el estudio de las ciencias sociales y el derecho, en Argentina y el mundo

Hace tiempo ya que la investigación en humanidades y ciencias sociales emplean algoritmos de inteligencia artificial para ofrecer perspectivas novedosas de los más variados temas. Por ejemplo, en *Mining the dispatch* se los utilizó para procesar casi la totalidad de las ediciones de un matutino del Estado de Virginia (EEUU) durante la Guerra Civil estadounidense, en busca de fluctuaciones en la cantidad de "anuncios de esclavos fugitivos", que el autor ofrece como una manera de sintetizar computacionalmente la historia de miles de hombres y mujeres que recuperaron su libertad durante esos años[25]. En *Gender bias in magazines oriented to men and women: a computational approach*, por otra parte, cientistas computacionales de Argentina identificaron los tópicos de una muestra de 8.000+ revistas para estudiar la diferencia en la que temas como "negocios", "ciencia", "familia", "moda", etc. son representados en medios locales, según si están orientados a varones o mujeres[26].

La inteligencia artificial también ha sido utilizada para estudiar diversos aspectos del derecho moderno, sobre todo en Europa y Estados Unidos, aunque lamentablemente su uso en Argentina es todavía embrionario. Parte del objetivo de este trabajo es, precisamente, remediar esa situación. Al promover la adopción de técnicas como las aquí descriptas en nuestro país, pretendemos en efecto acercar a la comunidad legal argentina al estado de la investigación en el mundo y, especialmente, inspirar futuros trabajos que contribuyan colectivamente al desarrollo de una academia jurídica computacional en nuestro país.

En *A General Approach for Predicting the Behavior of the Supreme Court of the United States*[27], por ejemplo, se utilizó un algoritmo de *machine learning supervisado* para predecir el comportamiento de la Corte Suprema de Estados Unidos. El modelo fue entrenado con 240.000 votos individuales de los jueces, y más de 28.000 votaciones de casos decididos por la Corte a lo largo de casi dos siglos (1816-2015), alcanzando una exactitud media del 70,2% en la predicción del sentido de las votaciones, y del 71,9% del sentido de los votos individuales.

De manera similar, en *Using Machine Learning to Predict Decisions of the European Court of Human Rights*[28], se utilizaron técnicas de aprendizaje automático para

---

[24] Ver MARTÍNEZ, E., "Re-Evaluating GPT-4's Bar Exam Performance", *SSRN*, 2023.

[25] NELSON, R.K., "Mining the Dispatch", University of Richmond, 2010.

[26] KOZLOWSKI, D., LOZANO, G., FELCHER, C.M., GONZALEZ, F., y ALTSZYLER, E.J., "Gender Bias in Magazines Oriented to Men and Women: A Computational Approach", ArXiv, 2020.

[27] KATZ, D.M., BOMMARITO, M.J., y BLACKMAN, J., "A General Approach for Predicting the Behavior of the Supreme Court of the United States", Plos One, vol. 12, 2017.

[28] MEDVEDEVA, M., VOLS, M., y WIELING, M., "Using Machine Learning to Predict Decisions of the European Court of Human Rights", Artificial Intelligence and Law, vol. 28, 2020.

predecir decisiones futuras del Tribunal Europeo de Derechos Humanos, mediante el procesamiento del texto de sus sentencias pasadas. El equipo detrás de la investigación reportó haber obtenido una exactitud promedio de 75% en la predicción de condenas por la violación de 9 artículos de la Convención Europea de Derechos Humanos.

A su turno, *Análisis Computacional del Derecho Argentino*[29] fue el primer trabajo de investigación que recurrió a algoritmos de inteligencia artificial para estudiar el derecho de nuestro país, mediante el procesamiento computacional del texto de más de 50.000 sentencias de la Cámara Federal de Casación Penal. Se utilizó una metodología similar a la descripta en los trabajos antes referenciados, pero con un propósito diferente: en lugar de intentar predecir decisiones futuras, el estudio se enfocó en la identificación de las propiedades de los casos que mejor pueden explicar los escenarios en los que la CFCP admitió o rechazó, en el pasado, los recursos que llegaron a sus estrados entre 2014 y 2021. En este caso, la exactitud alcanzada fue de 96%, en promedio[30].

Más cerca de la metodología que emplearemos para nuestra indagación sobre la jurisprudencia de la CFCP en esta presentación, mencionamos finalmente *On Top of Topics: Leveraging Topic Modeling to Study the Dynamic Case-Law of International Courts*[31] y *Reading the High Court at a Distance*[32], dos trabajos que aplicaron algoritmos de *topic modelling* para estudiar los temas principales que surgen de las sentencias de la Corte Europea de Justicia y el Tribunal Europeo de Derechos Humanos, por un lado, y de la Corte Suprema de Australia, por el otro. Los objetivos de esos trabajos no fueron idénticos: mientras que en el primero se buscó demostrar, de forma más o menos general, que estos algoritmos de IA pueden ser aplicados a la jurisprudencia de un tribunal, la investigación sobre precedentes australianos indagó en cuáles fueron las diez áreas de práctica legal —derecho de propiedad, proceso penal, inmigración, etc.— en las que ese tribunal superior había enfocado especialmente su atención a lo largo de más de dos siglos.

Para nuestra investigación utilizaremos una técnica similar para estudiar la jurisprudencia local. A diferencia de los trabajos referidos, que utilizaron muestras de 19.501 y 7.476 fallos, respectivamente, la muestra que analizaremos en esta

---

[29] MIELNIK, D., "Análisis Computacional del Derecho Argentino".

[30] Es importante destacar que este valor no es directamente comparable a los obtenidos por los estudios que abordaron la jurisprudencia de la Corte Suprema de Estados Unidos y del TEDH. No sólo porque se trata de tribunales con competencias disímiles sino, especialmente, porque los diferentes objetivos perseguidos en cada trabajo impactaron, naturalmente, en la metodología empleada en cada uno. Ver MIELNIK, D., "Análisis Computacional del Derecho Argentino", pp. 19–21.

[31] PANAGIS, Y., CHRISTENSEN, M.L., y SADL, U., "On Top of Topics: Leveraging Topic Modeling to Study the Dynamic Case-Law of International Courts", Frontiers in Artificial Intelligence and Applications, vol. 294, 2016.

[32] CARTER, D., BROWN, J., y RAHMANI, A., "Reading the High Court at a Distance: Topic Modelling the Legal Subject Matter and Judicial Activity the High Court of Australia, 1903–2015", The University of New South Wales Law Journal, vol. 39, 2016.

oportunidad contiene más de 80.000 sentencias, pronunciadas durante los 30 años que la CFCP lleva en funcionamiento. Este estudio, a su vez, es el primero en aplicar algoritmos de modelado de tópicos a jurisprudencia escrita en castellano y emitida por un tribunal argentino altamente especializado (*i.e.,* con competencia exclusivamente penal). Describimos la técnica con mayor detalle a continuación.

2. Modelado de tópicos latentes mediante IA

Hemos mencionado varias veces ya a los "algoritmos de modelado de tópicos", pero es momento de construir algunas primeras intuiciones acerca de esta técnica basada en IA, para poder entender en qué consiste exactamente y cómo puede ser utilizada para estudiar derecho.

En general, los seres humanos podemos inferir casi inmediatamente los temas principales a los cuales refiere un texto. Consideremos por ejemplo los siguientes fragmentos, tomados de la colección de precedentes analizados en esta investigación.

Fragmento n° 1:

> "[s]e le imputa haber integrado una asociación criminal destinada a ejecutar el plan criminal, sistemático y clandestino de represión estatal desarrollado en la República Argentina en el período 1976/1983, ordenado por los comandantes en jefe de las tres fuerzas armadas [...], consistente en la persecución de la población civil por motivos políticos, mediante la perpetración de detenciones ilegales de personas, su cautiverio clandestino en centros de detención en condiciones inhumanas, su interrogación bajo la aplicación de tormentos y, en muchos casos, su eliminación física"[33].

Fragmento n° 2

> "La política estatal, en vez de ubicar el problema de la droga como una cuestión sanitaria, lo transformó en un elemento de seguridad ciudadana, identificando a quien posee una adicción a los estupefacientes como un delincuente en vez de un droga-dependiente, pretendiendo crear, a través del derecho penal, un tipo de ciudadano modelo, sin que se respete la elección de su plan de vida. Esta prohibición ya la explicaba claramente John Stuart Mill en su célebre obra 'On Liberty', hace más de un siglo"[34].

En efecto, no nos lleva más que unos segundos advertir que los temas centrales del primer fragmento están asociados a los crímenes contra la humanidad, la violación de derechos humanos, la dictadura militar y la represión; y que los temas del segundo se vinculan con el tratamiento punitivo del consumo de estupefacientes, la autonomía personal y la filosofía liberal. Seguramente, incluso acordaremos en que no es necesario siquiera leer los fragmentos completos para inferir sus temas, y quizás nos

[33] Sala IV CFCP, "San Martín y otros" (causa FGR 33008736/2005/TO4/CFC12, reg. n° 727/21.4, del 2/6/2021).

[34] Sala III CFCP, "Guerra, Jorge Néstor Rubén" (causa n° 6472, reg. n° 128/2006.3, del 7/3/2006).

sorprenda verificar que las palabras "dictadura" o "liberal" ni siquiera aparecen en ellos.

Para efectuar esta operación, no hemos hecho más que observar, quizás inconscientemente, un puñado de palabras que, aislada o conjuntamente, asociamos al campo semántico de cada uno de los temas que nos vienen a la mente. En este sentido, "persecución de la población civil", "comandantes en jefe de las tres fuerzas armadas" y "centro clandestino de detención" son palabras y expresiones asociadas a los crímenes contra la humanidad perpetrados durante la última dictadura que gobernó el país; mientras que "estupefacientes", "elección del plan de vida" y "John Stuart Mill" son parte del campo semántico en el que discurre el debate sobre la criminalización de la tenencia de estupefacientes para consumo personal y su inserción en el derecho penal de un Estado liberal.

Si repetimos esta operación para todos los documentos de una colección, al final obtendremos dos cosas: una lista con los términos más característicos de cada tema (o tópico) presente en la colección, y una de los temas presentes en cada documento. Eso es en esencia un modelo de tópicos, al menos en lo que puede ser de interés, en principio, para un/a jurista: una representación de la manera en la que ciertas palabras y expresiones tienden a ocurrir conjuntamente —y por lo tanto hacen a un tópico en particular— y de la manera en la que los tópicos de la colección se combinan y permiten describir cada documento en particular. Desde las humanidades digitales, el proceso ha sido descripto de la siguiente manera:

> "Una forma de pensar acerca del proceso de *topic modelling* es imaginar que vamos leyendo un artículo con un conjunto de resaltadores de colores. A medida que avanzamos en la lectura, utilizamos un resaltador diferente para identificar las palabras clave de un tema dado. Cuando terminamos, copiamos todas las palabras de un mismo color y les asignamos una etiqueta. La lista de palabras constituye el tópico, y cada color representa un tópico diferente"[35].

Los algoritmos de modelado de tópicos o *topic modelling* intentan emular esta clase de operaciones mediante mecanismos informáticos automatizados. El que utilizaremos para esta investigación es el, así llamado, algoritmo de *Asignación de la Dirichlet Latente*, o LDA, por las siglas de su nombre en inglés: *Latent Dirichlet Allocation*[36]. Todo lo que decimos sobre el modelado de tópicos en este trabajo debe entenderse en relación con esta técnica —que por cierto no es la única, aunque sí, y por lejos, la más popular—.

A diferencia de nosotros, LDA no tiene acceso al contexto y los saberes previos que nos permiten relacionar expresiones como "engaño que produce una lesión

---

[35] BRETT, M.R., "Topic Modeling: A Basic Introduction", Journal of Digital Humanities, vol. 2, 2012.

[36] BLEI, D.M., NG, A.Y., y JORDAN, M.I., "Latent Dirichlet Allocation", Journal of Machine Learning Research, vol. 3, 2003.

patrimonial" y "valoración de agravantes y atenuantes" a los tópicos "estafa" y "graduación de la pena", respectivamente. En su lugar, el algoritmo es capaz de inferir esas relaciones directamente a partir de la manera en que las palabras ocurren dentro de cada documento, y a lo largo de toda la colección[37]. En efecto, LDA es un algoritmo probabilístico y generativo, en el sentido de que asume, como presupuestos, que cada documento del *corpus* se corresponde con una combinación particular de tópicos, y que cada tópico, a su vez, es una combinación particular de términos y expresiones.

Así, por ejemplo, LDA entendería un hipotético fallo sobre la admisibilidad del recurso contra decisiones que prorrogan la prisión preventiva en casos de estafa como un documento generado en un 50% por el tópico "prisión preventiva", 30% por expresiones asociadas al tópico "admisibilidad del recurso de casación", 20% por el tópico "estafa", y 0% para tópicos como "homicidio" o "valoración de la prueba"[38]. A su vez, describiría el primer tópico como una combinación de términos tales como "prisión preventiva", "prórroga", "riesgo procesal", etc.; el segundo, como una en la que "recurso formalmente admisible", "impugnabilidad objetiva", "sentencia equiparable a definitiva", "artículo 457 del C.P.P.N.", etc. tienen mayor probabilidad de ocurrencia; y el tercero, como una en la que"fraude", "ardid", "engaño", "lesión patrimonial", etc. tienen especial preponderancia[39]

Esos presupuestos son obviamente una simplificación de la realidad, pero en todo caso están anclados en la observación de que, en un sentido genuino, hablar de estafas o escribir un artículo académico sobre la dogmática penal implica, en alguna medida, utilizar un cierto conjunto de palabras y expresiones características del discurso respectivo. Como mínimo, es muy difícil hacerlo sin recurrir a esas expresiones.

Por cierto, aunque el nombre del algoritmo puede sonar algo críptico, la expresión *latent allocation* hace referencia, precisamente, a la capacidad de este sistema de IA para asignar cada combinación particular de palabras que suelen co-ocurrir a un tópico determinado, y cada documento de una colección a una combinación particular de tópicos, que antes del proceso se encontraban ocultos o latentes. Sin entrar en tecnicismos, *Dirichlet*, por su parte, es el nombre de un recurso estadístico que puede utilizarse para describir tanto la forma en la que *a priori* el algoritmo espera que las palabras de la colección queden asociadas a un cierto tópico, como la manera en la que un conjunto de tópicos dado quede asociado a cada documento[40].

Así, LDA es en efecto capaz de aprender por sí mismo cuáles son los tópicos presentes en una colección de textos arbitrariamente extensa, y cuáles de ellos

---

[37] BLEI, D.M. *et al.*, "Latent Dirichlet Allocation", p. 996.

[38] Dada la naturaleza probabilística de LDA, sin embargo, en la práctica ningún tópico es asignado 0% de probabilidad, sino un valor tan bajo que resulta obviamente despreciable.

[39] Exhibimos los tópicos efectivamente aprendidos de nuestro *corpus* de fallos en la Sección 3.

[40] La *Dirichlet* posee además ciertas propiedades matemáticas que facilitan las computaciones necesarias para que el sistema vaya ajustando sus parámetros de funcionamiento interno de manera eficiente.

probablemente contribuyeron en mayor medida a la generación de cada documento. De esa manera, si lo aplicamos a un *corpus* compuesto por fallos de un tribunal penal —como es nuestra intención hacer—, esperaríamos que el algoritmo aprenda que los términos "estupefacientes", "droga" y "ley 23.737" tienden a aparecer juntos —en los mismos documentos— y que en esos documentos, a su vez, expresiones como "falsedad ideológica", "firma apócrifa", "peritaje caligráfico" sólo aparecerán en una proporción mucho menor (si es que aparecen), al tiempo que sí tienden a co-ocurrir en otros artículos (precisamente, en los que el tópico vinculado a los delitos de falsedad documental tenga mayor preponderancia).

Es importante señalar que, dado que cada tópico está definido como una combinación particular de las probabilidades con las que que ciertos términos se encuentren en un documento dado, la asignación de la Dirichlet latente no es técnicamente un proceso de *clustering*, ya que distintos tópicos pueden compartir términos en común. Así, por ejemplo, el término "tenencia" puede aparecer con alta probabilidad tanto en un tópico sobre tenencia de estupefacientes como de armas de fuego, y el término "salud" puede tener representación tanto en un tópico sobre la pandemia de COVID-19 como uno referido a las condiciones para la incorporación al régimen de prisión domiciliaria (art. 10, inc. "a" CP).

El proceso, por cierto, no es completamente autónomo. En primer lugar, si bien LDA aprende a ajustar sus parámetros de funcionamiento interno de manera automática —eso lo convierte en un algoritmo de IA, en el sentido más moderno precisado en la sección 1— posee también algunos *hiperparámetros* que debemos ajustar manualmente; algo así como las perillas de nuestro robot, que afectan los presupuestos con los que se acerca a la muestra que procesará.

Algunos de esos hiperparámetros, como los habitualmente llamados α y β, representan las expectativas *a priori* que tenemos sobre la manera en la que los tópicos se presentarán a lo largo de la colección: con ellos le comunicamos a LDA si esperamos que los documentos del *corpus* estarán compuestos por muchos tópicos o sólo por algunos, y si los tópicos en sí contendrán mixturas más o menos homogéneas de términos co-ocurrentes. El hiperparámetro más importante, sin embargo, es la cantidad *k* de tópicos que pretendemos identificar algorítmicamente, aunque es importante recordar que, al tratarse de un proceso *no supervisado*, no hay un número objetivamente correcto de tópicos, sino que éste depende de las necesidades del caso de uso concreto[41] y del tamaño de la muestra.

En general, un número pequeño de tópicos inducirá al algoritmo a producir representaciones temáticas más generales y abarcativas, mientras que uno grande producirá modelos más granulares y matizados. Sin embargo, la calidad del modelo producido puede resultar afectada tanto por exceso como por defecto: una *k* demasiado pequeña hará colapsar, en un mismo tópico, términos que un ser humano quizás consideraría propios de dos o más temas diferentes; por otro lado, si la cantidad

[41] Ver CARTER, D. *et al.*, "Reading the High Court at a Distance", p. 1311.

pretendida es excesivamente grande, el sistema terminará generando artefactos ininteligibles, compuestos por fluctuaciones estadísticas sin carga semántica genuina[42].

Por lo demás, hay que señalar que LDA puede identificar los grupos de palabras que conforman tópicos, pero es incapaz de etiquetarlos por sí mismo. Así, será nuestra responsabilidad denominar "Derecho Civil" al tópico compuesto, por ejemplo, por los términos "contrato", "demanda", "reconvención", "daños y perjuicios", etc., y "Derecho laboral" al tópico que aglutina expresiones como "despido", "licencia sin goce de sueldo", "convenio colectivo", etc.

Estos aspectos resaltan una vez más que los algoritmos de inteligencia artificial, por asombrosos que resulten, no dejan de ser herramientas. Así, pueden completar en segundos o minutos tareas que, como la identificación de los temas presentes en cientos de miles de fallos judiciales, a los seres humanos nos insumirían varios años (tal vez varias vidas) y, por ese motivo, nos estarían vedadas en la práctica. Pero el proceso y sus resultados requieren que los interpretemos y los pongamos en la perspectiva que sólo el dominio sobre la materia estudiada nos puede dar.

3. Modelado de tópicos y jurisprudencia argentina

Es momento de completar nuestra aproximación a los usos de la inteligencia artificial para el estudio del derecho con una aplicación práctica de las técnicas de procesamiento computacional mediante IA discutidas en las secciones anteriores. Nuestro propósito en esta sección, en efecto, es demostrar la utilización de un algoritmo de *topic modelling* para estudiar la jurisprudencia nacional a una escala nunca antes alcanzada. En particular, analizaremos una muestra de fallos de la Cámara Federal de Casación Penal —la máxima instancia del ordenamiento penal federal— pronunciados a lo largo de 30 años, entre diciembre de 1992 y diciembre de 2022. La técnica elegida, así, permitirá efectuar un estudio exploratorio y retrospectivo que abarcará prácticamente toda la jurisprudencia digitalmente publicada de la Cámara.

3.1 La muestra analizada

Hay varias razones para centrar la atención en la CFCP en estos primeros estudios computacionales de la jurisprudencia argentina[43]. Primero, la competencia de la CFCP se extiende a todo el territorio argentino, por lo que estudiar su jurisprudencia es en parte estudiar la de cada uno de los distritos judiciales que componen la jurisdicción

[42] Estas particularidades del algoritmo LDA quedarán más claras con la aplicación práctica que presentamos en la sección siguiente.

[43] MIELNIK, D., "Análisis Computacional del Derecho Argentino", pp. 13–14.

federal y, a través de ellos, efectuar una lectura a distancia de lo que sucede en los tribunales con competencia penal a lo largo y ancho del país.

A su vez, a partir del precedente de la Corte Suprema "Giroldi"[44], y especialmente luego de "Casal"[45], la CFCP se convirtió en el tribunal federal responsable de garantizar el derecho a impugnar el fallo condenatorio, reconocido en los arts. 8.2 "h" de la Convención Americana sobre Derechos Humanos y 14.5 del Pacto Internacional de Derechos Civiles y Políticos, que exige la revisión integral de los casos que llegan a su conocimiento —incluyendo a las denominadas "cuestiones de hecho y prueba"—, hasta agotar sus posibilidades de rendimiento revisorio. En lo que importa para esta investigación, eso significa que las decisiones estudiadas contienen en general una síntesis de todo el proceso penal que condujo a su dictado, lo que también enriquece potencialmente el análisis y sus conclusiones.

La muestra analizada está compuesta íntegramente por fallos publicados en formato digital y son de acceso libre. En efecto, desde 2013, la normativa vigente garantiza que todas las decisiones de la Corte Suprema, las cámaras federales y nacionales de apelación y casación, así como las sentencias definitivas de los tribunales orales, sean publicadas el mismo día de su pronunciamiento. A su turno, la ley 27.275 (2017) estableció el carácter de *información pública* de la jurisprudencia de esos tribunales, e incluso ordena ponerla a disposición de la ciudadanía sin necesidad de un pedido expreso de acceso a la colección, bajo el denominado "principio de transparencia activa". La Figura 1 resume cómo se distribuyen los fallos de la muestra a lo largo de los años.

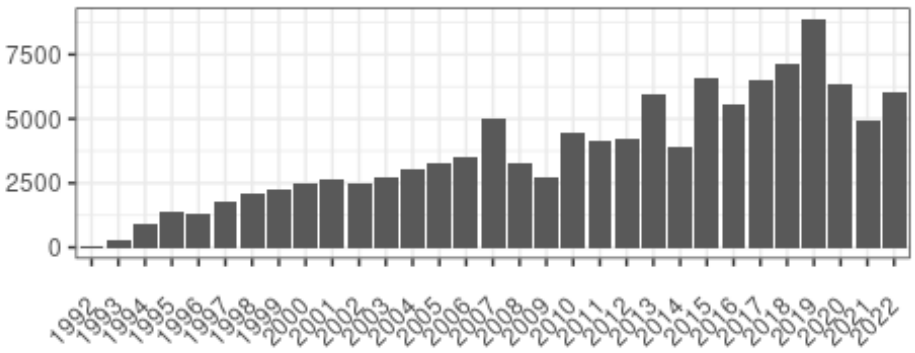


Figura 1: Cantidad de fallos por año incluidos en la muestra

Antes de entrenar el modelo de tópicos, es habitual realizar una serie de tareas de limpieza y pre-procesamiento para mejorar la calidad de datos que se ingresan en el modelo. Para nuestro estudio, por ejemplo, comenzamos eliminando los artefactos

[44] Fallos: 318:514.

[45] Fallos: 328:3399.

generados durante el proceso de digitalización de las sentencias[46]. También aplicamos un método estadístico[47] para encontrar expresiones compuestas —llamadas *collocations,* en lingüística computacional— que poseen significados diferentes que la de sus partes, tomadas aisladamente (por ejemplo, "violencia de género" posee un significado diferente que "violencia" y "género"), de modo que el algoritmo de *topic modelling* las considere como un solo término. Por otra parte, aplicamos un filtro sintáctico con el objeto de preservar únicamente los sustantivos y los adjetivos del texto, que fueron considerados las clases de palabras con mayor poder descriptivo para la inferencia de tópicos jurisprudenciales. Finalmente, eliminamos del texto todas las palabras con menos de 3 letras, así como las denominadas *stopwords*: conectores y otras palabras ubicuas ("de", "que", "entonces", "máxime", etc.), que tampoco poseen información relevante para un modelo de tópicos como el que pretendemos generar mediante IA.

3.2 Los tópicos de casación penal

Para este trabajo configuramos LDA para inferir un modelo de 100 tópicos que, dado el gran tamaño de la muestra, permitió alcanzar razonable granularidad e interpretabilidad de los temas latentes en la muestra[48]. Por supuesto, lo primero que nos interesa conocer son los temas que el algoritmo aprendió a identificar. Si bien en las páginas siguientes los iremos analizando uno a uno (en la medida en que resulten útiles), la Tabla 1 muestra una selección de algunos de ellos[49].

Tabla 1: Selección de tópicos del modelo generado, junto con las 10 palabras de mayor probabilidad asociada a cada uno y la descripción (etiqueta) asignada por los autores.

ID	Términos	Descripción
6	domiciliaria, salud, arresto_domiciliario, medicos, patologias, medico, prision, detencion, medica, alojamiento	Prisión domiciliaria por razones médicas
8	violencia, mujer, mujeres, genero, violencia_de_genero, discriminacion, varones, sexual, formas, politicas_publicas	Violencia de género
14	pandemia, salud, higiene, medidas_de_prevencion, feria, guia, emergencia_sanitaria, acordada, prevencion, sanitaria	Pandemia
28	victima, lesiones, homicidio, muerte, agresion, disparos, damnificado, herida, cuchillo, golpes	Homicidio
39	derechos, humanos, internacional, argentino, crimenes, violaciones, responsables, obligacion, humanidad, comunidad_internacional	Violaciones de derechos humanos

[46] Por ejemplo, símbolos y caracteres extraños al idioma castellano, o números de serie incorporados por el sistema informático de gestión de expedientes utilizado por el Poder Judicial de la Nación.

[47] CHURCH, K.W. Y HANKS, P., "Word Association Norms, Mutual Information, and Lexicography", Proceedings of the 27th Annual Meeting on Association for Computational Linguistics, 1989.

[48] El rendimiento del modelo fue evaluado a través de la observación empírica de los tópicos generados y mediante el cálculo de la *coherencia CV* media de los principales términos que definen cada tópico, esto es, en esencia, el grado de cercanía semántica entre ellos. Ver RÖDER, M., BOTH, A., y HINNEBURG, A., "Exploring the Space of Topic Coherence Measures", *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, ACM, Shanghai China, 2015.

[49] La tabla completa se puede consultar en https://dijkie85.github.io/topicoscfcp/topicos.html.

ID	Términos	Descripción
40	arma, robo, arma_de_fuego, armas, banda, calibre, disparo, pistola, desapoderamiento, damnificado	Robo
41	reduccion, estimulo_educativo, libertad, fases, asistida, institutos, periodo, temporales, periodos, plazos	Estímulo educativo
62	niño, ninos, menores, madre, hijos, domiciliaria, familiar, cuidado, interes, superior	Prisión domiciliaria por cuidado de menor
65	agravada, privacion, ilegal, tormentos, concurso, humanidad, amenazas, victimas, libertad, imposicion_de_tormentos	Violaciones de derechos humanos
75	cion, i, fos, a, o, r, t, miento, l, e	Artefacto estadístico
82	estafa, contrato, patrimonial, inmueble, quiebra, patrimonio, escritura, firma, ardid, empresa	Estafas y defraudaciones
84	prueba, testigo, dichos, testigos, testimonio, testimonios, certeza, declaraciones, pruebas, valoracion	Prueba testimonial

Elegimos estos tópicos para ilustrar algunos aspectos del algoritmo de *topic modelling* que adelantamos en la Sección 2. Para empezar, aquí se ve con claridad que los tópicos inferidos mediante IA incluyen los temas de derecho sustantivo a los que suele hacer referencia la expresión "tópico" en la discusión jurídico-penal (delitos contra la propiedad [tópicos 40 y 82], delitos de lesa humanidad [tópicos 39 y 65], etc.), pero el sentido capturado por LDA es también mucho más amplio —en efecto, el algoritmo identifica cualquier tema abordado de manera más o menos estable en la muestra de jurisprudencia—. Esa propiedad por sí sola puede tener un valor inmenso para la enseñanza del derecho, al permitirnos formular, más allá de los temas discutidos en los libros de teoría, una "taxonomía de la práctica"[50]: el conjunto de discusiones que un tribunal dado aborda *en la realidad* y que, por esa razón, se podría argumentar que constituyen los contenidos mínimos que cualquier estudiante o aspirante a ingresar al Poder Judicial debería dominar.

La selección muestra también que el modelo posee cierta granularidad y matices: dentro del tema más general del "Arresto domiciliario", por ejemplo, LDA fue capaz de distinguir según si el motivo que lo fundamenta es el deterioro de la salud de la persona detenida (tópico 5; art. 10, inc. "a" CP) o el cuidado de un/a menor (tópico 61; art. 10, inc. "f" CP). Dependiendo de los propósitos de la investigación, sin embargo, en algunos casos esa misma granularidad puede resultarnos excesiva. Por ejemplo, el tópico 39 está compuesto principalmente por términos vinculados a la obligación de prevenir, investigar y juzgar graves violaciones a los derechos humanos, mientras que el tópico 65 recoge expresiones asociadas a las figuras penales que esos crímenes suelen entrañar en particular (privación de la libertad agravada, tortura, desaparición forzada, entre otras). Así, si bien no hay nada incorrecto en la inferencia algorítmica —se trata en efecto de dos discursos distintos— alguien podría señalar que, si la tarea de identificación de tópicos hubiera sido encargada a un ser humano, en lugar de una IA, éste probablemente habría anotado bajo un mismo rótulo ("Violaciones de derechos humanos") tanto a los fallos con alta presencia del tópico 39, como aquellos en los que el tópico 65 es predominante.

[50] CARTER, D. *et al.*, "Reading the High Court at a Distance", p. 1338.

Este tipo de resultados, por cierto, no es problemático: dado que LDA modela la presencia de los tópicos en función de la probabilidad de que los encontremos a lo largo de cada documento, los tópicos que consideremos fragmentados pueden agruparse en un tópico unificador, que puede modelarse a partir de la suma de las probabilidad de los temas subyacentes, dado que $P(t_1 + t_2) = P(t_1) + P(t_2)$. Propiedades similares del modelo de tópicos permiten también que podamos descartar aquellos que, como el tópico 75, están conformados básicamente por artefactos estadísticos, o que, por la razón que fuere, resulten inútiles o ininteligibles.

Por lo demás, podemos validar el modelo haciendo una "prueba de sanidad", mediante la observación de algún tópico cuyo comportamiento resulte especialmente evidente. La Figura 2 muestra, por ejemplo, qué sucede con el tópico "Pandemia" a lo largo del tiempo. Como uno esperaría, el tema carecía casi por completo de presencia antes de 2020, luego se incrementa vertiginosamente y finalmente va mermando a medida que evoluciona la situación sanitaria en el país. Eso evidencia la capacidad del algoritmo para realizar su trabajo de inferencia correctamente.

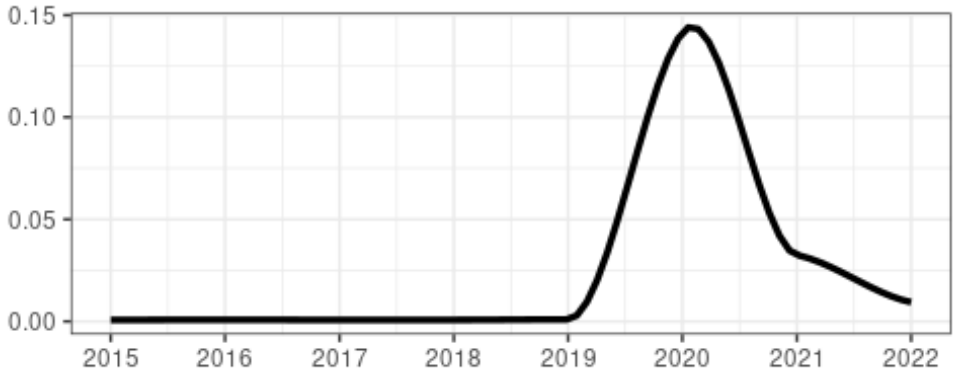


Figura 2: Evolución del tópico "Pandemia" desde el 2015 en adelante

3.3 Tópicos más prevalentes

Miremos más de cerca la distribución de tópicos de nuestro modelo. Nos puede interesar, por ejemplo, utilizarlo para identificar los principales temas discutidos en los fallos que componen el *corpus*. La Figura 3 muestra la presencia relativa de los 10 principales temas abordados en la jurisprudencia analizada y pone en evidencia una circunstancia quizás inesperada para quien no está acostumbrado a litigar ante este tribunal: que, en efecto, la mayor parte de los fallos de la CFCP está dedicada a resolver cuestiones más bien formales o de escaso valor nomofiláctico, como la admisibilidad de los distintos recursos que llegan al tribunal, o la resolución de conflictos de competencia suscitados en instancias inferiores.

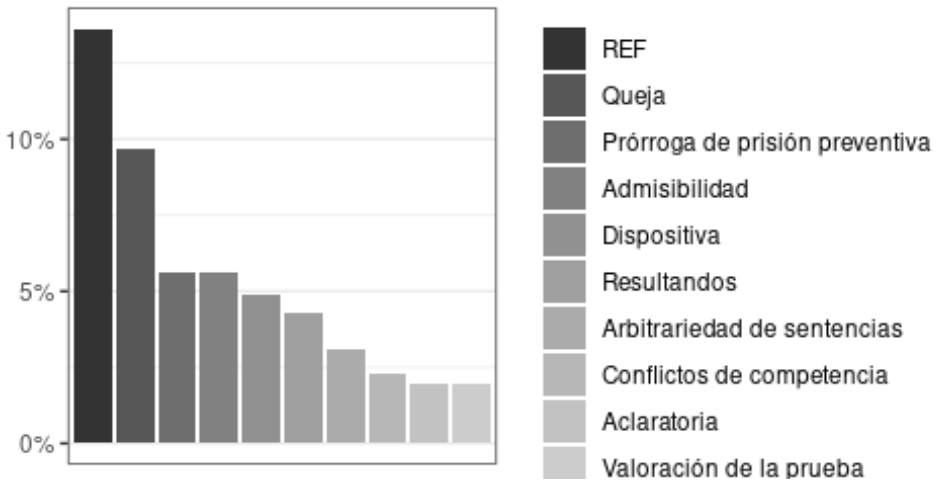


Figura 3: Principales tópicos en el corpus

Es cierto, sin embargo, que tópicos como "Parte dispositiva" o "Resultandos" probablemente suman prevalencia por el hecho de que aparecen en casi la totalidad de los fallos. Para tener entonces una mejor perspectiva del lugar que ocupa la resolución de cuestiones formales en el trajín diario de la CFCP podemos, por ejemplo, sumar las probabilidades de los tópicos correspondientes, y calcular la proporción de fallos en los que esos temas constituyen la mayor parte del texto (es decir, los casos en los que las cuestiones formales son el principal tema discutido). De esta manera, observamos que aproximadamente en 48% de los casos resueltos por la CFCP, los tópicos formales representan más del 50% del contenido de esos fallos. La Figura 4 muestra cómo cambia esa proporción a medida que variamos el umbral por encima del cual consideraremos que el fallo trata centralmente cuestiones no sustantivas.

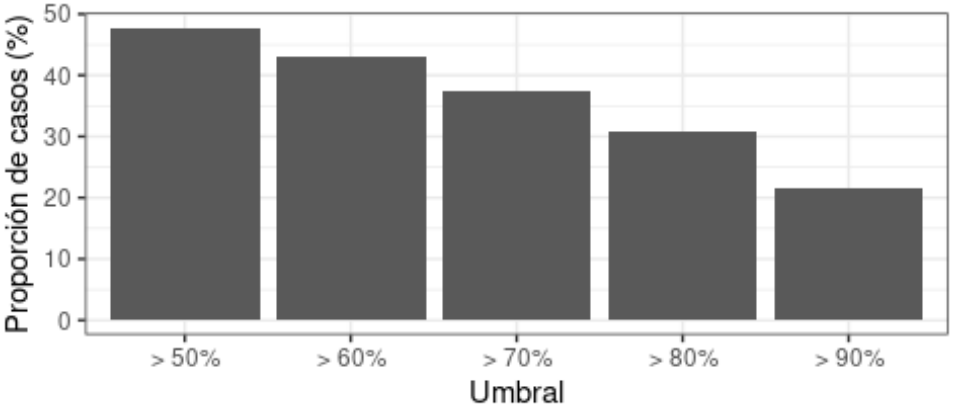


Figura 4: Proporción de casos en los que los tópicos formales son el contenido principal para distintos valores de umbral.

Así, el estudio de la prevalencia de tópicos formales permite estimar que, dependiendo de donde coloquemos el umbral, la CFCP dedica 21,46%-47,82% de su capacidad a resolver casos en los que la cuestión central es en algún sentido

puramente formal, lo que puede sugerir la existencia de problemas sistémicos en el proceso recursivo que, así, la IA nos ayuda a diagnosticar.

3.4 Los tópicos sustantivos de la parte especial

Descontemos ahora los tópicos puramente formales y centrémonos, en particular, en los tópicos de la parte especial del derecho penal sustantivo: las clases de delitos litigados ante la CFCP. El modelo generado logró identificar las siguientes: "homicidio", "robo", "robo / secuestro extorsivo", "abuso sexual", "estafas y defraudaciones", "homicidio y lesiones imprudentes", "estupefacientes (general)", "estupefacientes (consumo personal)", "violaciones de derechos humanos", "estupefacientes (transporte)", "estupefacientes (comercio)", "falsedad documental", "delitos tributarios", "delitos tributarios (ley más benigna)", "trata y explotación de personas", "delitos de funcionarios públicos", y "delitos aduaneros", que ciertamente parecen abarcar la mayor parte de los ilícitos que, en efecto, esperaríamos que lleguen a conocimiento de este tribunal.

Para precisar un poco más el análisis, la Figura 5 muestra cuáles fueron los delitos que más representación tuvieron en la jurisprudencia de los primeros años de la CFCP —digamos, de 1992 a 2005—, y cuáles ocupan los primeros lugares desde 2016.

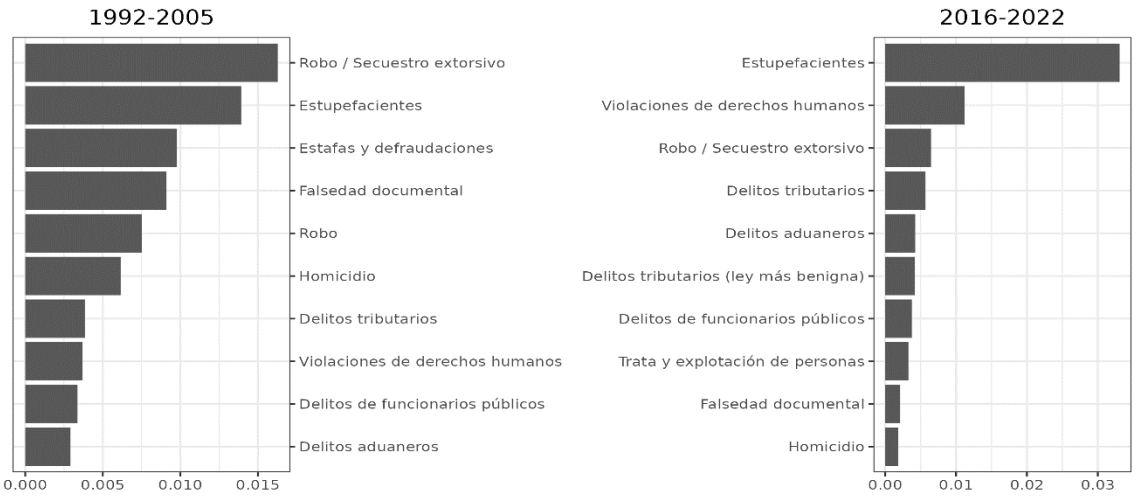


Figura 5: Principales tópicos sustantivos

Lo que tal vez llame la atención aquí es la virtual ausencia de delitos de competencia ordinaria en el período 2016-2022. No hay que sacar conclusiones apresuradas: lo que probablemente explique la Figura es que esos ilícitos pasaron en 2015 a la órbita de la Cámara Nacional en lo Criminal y Correccional, por lo que es esperable no entrarlos en decisiones recientes. La Figura 6 muestra con claridad el descenso de los casos vinculados con esas temáticas a partir de 2015[51].

[51] Por cierto, el tópico "Robo / Secuestro extorsivo" es algo ambiguo en este caso ya que recoge delitos pertenecientes a distintos fueros. La conexión identificada por LDA se explica probablemente en razón

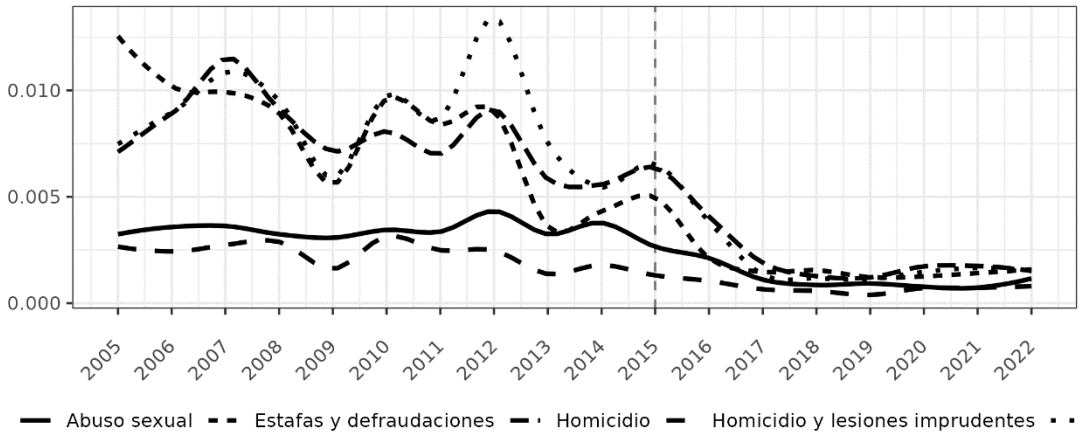


Figura 6: Presencia de delitos ordinarios a lo largo del tiempo

3.5 Violaciones de derechos humanos y violencia de género

Detengámonos un momento para analizar el comportamiento de dos tópicos que han recibido una enorme atención en los últimos años y que, en alguna medida, reflejan la forma en que lidiamos, como sociedad, con dos aspectos dramáticos de nuestra historia y vida en comunidad: las graves violaciones a los derechos humanos perpetradas durante la última dictadura (1976-1983) y la violencia de género.

La Figura 7 muestra cómo evolucionó el juzgamiento de los crímenes contra la humanidad en la jurisprudencia de la CFCP. En ella aparece cifrada una parte importante de la historia argentina: clausurada su investigación por las leyes de obediencia debida y punto final[52], primero, y por los indultos a los represores, después, los juicios pudieron retomarse recién en 2003, con la sanción de la ley 25.779 y, especialmente, con el dictado de diversos precedentes de la Corte Suprema que la respaldaron y allanaron el camino para su reapertura. Así, la curva muestra claramente cómo creció la presencia de la temática a partir de 2004, acompañando los sucesivos precedentes de la CSJN sobre la materia: "Arancibia Clavel" (2004)[53], "Simón" (2005)[54], "Mazzeo" (2007)[55], "Videla" (2010)[56], etc. El tópico alcanzó también un pico de prevalencia en 2017, año en que la CSJN consideró aplicable el cómputo privilegiado de días de prisión preventiva (2x1) para condenados por su intervención

de la relativamente escasa cantidad de sentencias sobre secuestros, en relación con los robos, durante la mayor parte de la vida de la CFCP, y por la cercanía semántica del discurso que los involucra (*i.e.*, uso de vehículos, armas, violencia, desapoderamiento, etc.).

[52] Leyes n° 23.521 y 23.492.

[53] Fallos: 327:3294.

[54] Fallos: 328:205.

[55] Fallos: 330:3248.

[56] Fallos: 333:1657.

en la represión ilegal, propiciando una catarata de planteos con pretensiones similares ante la CFCP que —hay que señalarlo— ésta invariablemente rechazó.

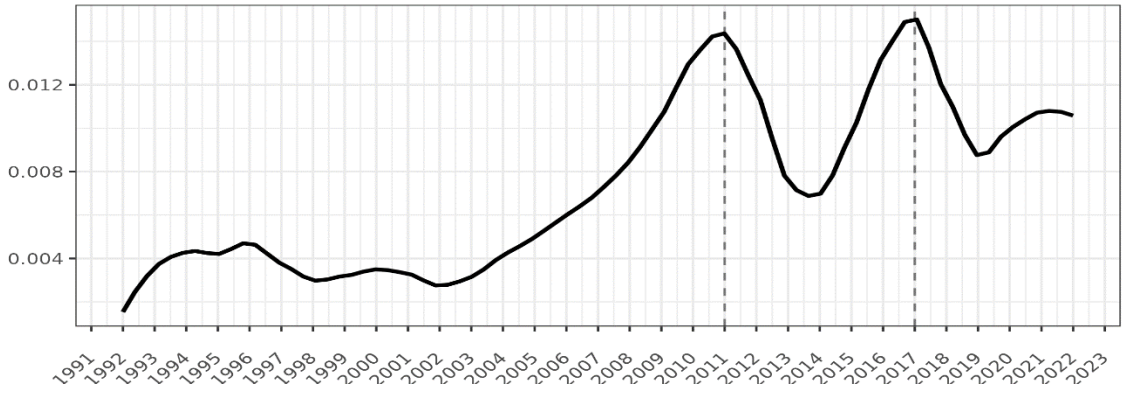


Figura 7: Violaciones de derechos humanos

El análisis del tópico "Violencia de género" (Figura 8) también muestra un patrón interesante: sostenidamente ausente durante la mayor parte de la historia de la CFCP, evidencia un crecimiento abrupto a partir de 2010-2011 y toca su pico en 2014. En este sentido, se puede recordar que en 2009 se sancionó la ley 26.485, de "Protección integral para prevenir, sancionar y erradicar la violencia contra las mujeres en los ámbitos en que desarrollen sus relaciones interpersonales", y en ese mismo año se creó la Oficina de la Mujer, en el ámbito de la CSJN. Fueron también los años en los que las mujeres pusieron el tema en la agenda pública como nunca antes, que pasó a ocupar un lugar central y definitivo a partir de 2015, con la marcha #NiUnaMenos.

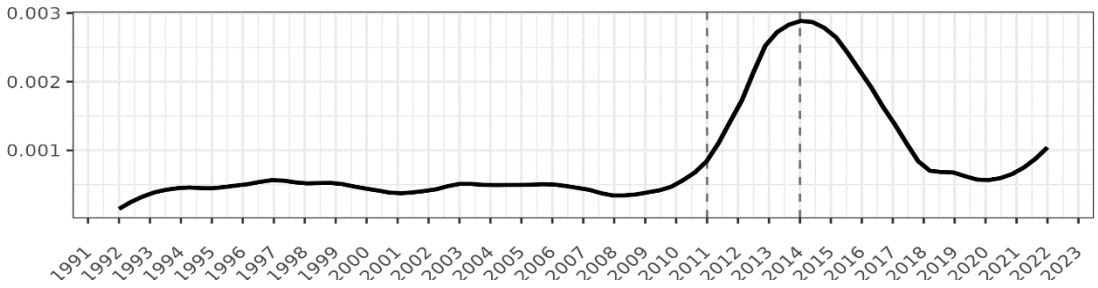


Figura 8: Violencia de género (CFCP)

El descenso, también pronunciado, probablemente se explique (al menos en parte) en razón de que los principales delitos en los que se juega la perspectiva de género (abuso sexual, femicidio, amenazas, lesiones, etc.) fueron retirados de la competencia de la CFCP en 2015, como recordamos más arriba (Figura 6). Los datos son consistentes también con investigaciones previas, que mostraron un patrón similar en

un análisis de 18.170 sentencias de los Tribunales Orales de la Ciudad de Buenos Aires, pronunciadas entre 2014 y 2019[57]. Reproducimos aquí esa figura (Figura 9).

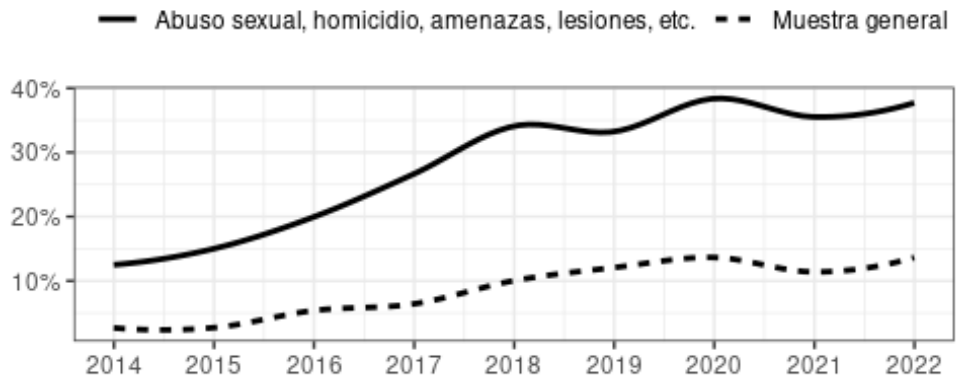


Figura 9: Violencia de género en TOC (Tomado de Mielnik, D., "Ciencia de datos y análisis del derecho penal", p. 183. La figura muestra el crecimiento del tópico al menos desde 2014, y diferencia entre una muestra general, y una compuesta sólo por fallos en los que el delito pertenecía a una selección en los que la mirada de género es especialmente urgente.)

También en el marco de la perspectiva de género en el ámbito judicial, podemos observar cómo se comporta el tópico "Acusada mujer" en relación con los tópicos vinculados con clases de delitos relevados más arriba (Figura 5). De esa manera, por ejemplo, es posible estimar la proporción en la que distintas clases de ilícitos tienen como imputada a una mujer. Como muestra la Figura 10, esa proporción es especialmente elevada en casos de trata de persona (~13%), esto es, una clase de ilícitos en los que las víctimas también son, desproporcionadamente, mujeres. El estudio de jurisprudencia mediante IA nos permite, así, investigar empíricamente ciertos fenómenos que a veces es muy difícil medir a gran escala. En este caso, el análisis parece respaldar la existencia del dramático fenómeno —descripto en la literatura— de mujeres víctimas de trata que se convierten a su vez en tratantes de otras mujeres[58].

---

[57] Ver MIELNIK, D., "Ciencia de Datos y Análisis Del Derecho Penal", p. 183.

[58] HOPP, C., "La lucha contra la trata de mujeres: la criminalización de las 'malas víctimas' como consecuencia no deseada", en JARAMILLO, I. y CORREA, M.C. (eds), *Sexo, violencia y castigo*, Didot, 2020. Hay que notar que si bien el análisis automatizado mediante IA revela una relación similar entre trata de personas y acusadas mujeres, la proporción encontrada es menor que la reportada en estudios previos: los datos de la unidad fiscal especializada en trata de personas (PROTEX), por ejemplo, documentan hasta un 37% de mujeres condenadas por trata. La diferencia puede explicarse por varios motivos. Por un lado, el informe de PROTEX (https://www.mpf.gob.ar/protex/recurso/informe-2014-sobre-las-primeras-100-sentencias-condenatorias-por-trata-de-personas/) recoge datos de 220 personas condenadas en primer instancia entre 2009 y 2014, mientras que nuestro análisis releva 5.044 decisiones interlocutorias y definitivas dictadas por la CFCP entre 1992 y 2022, en los que LDA identificó una o más mujeres acusadas. En cualquier caso, la diferencia observada sirve para poner una nota de cautela sobre las formas de análisis automático y la necesidad de validación de sus resultados.

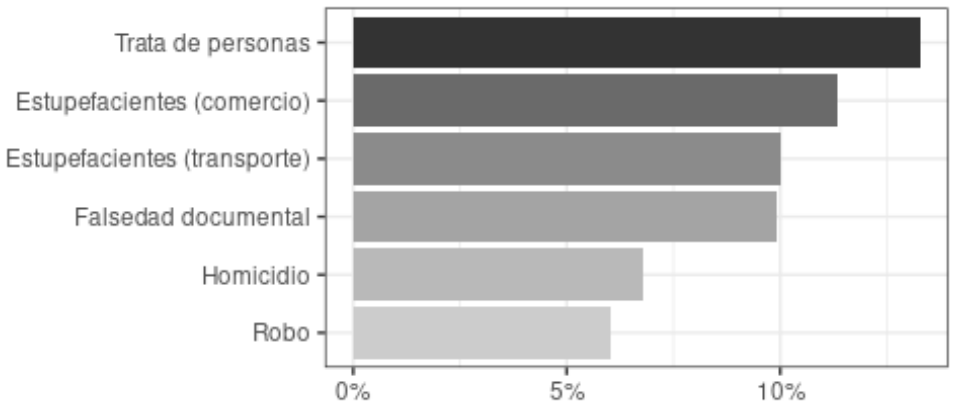


Figura 10: Proporción de acusadas mujeres por clase de delito

3.6 Relación de la CFCP con la Corte Suprema y consigo misma

Completamos este recorrido por el análisis algorítmico de 30 años de jurisprudencia de casación penal con algunas observaciones sobre la relación entre la CFCP y su tribunal de apelación —la Corte Suprema de Justicia de la Nación—, y consigo misma. Tomemos como ejemplo de la primera relación el comportamiento del tópico vinculado al alcance que debe darse a la revisión de las sentencias recurridas en casación (Figura 11).

Originalmente, la CFCP fue concebida como un tribunal nomofiláctico de intervención más bien excepcional, limitada a la interpretación del derecho penal sustantivo y el control jerárquico-burocrático del proceso penal. Eso comenzó a cambiar en 1995, con el dictado del precedente "Giroldi", en el que la CSJN interpretó que el derecho convencional a la revisión del fallo condenatorio[59] impide que aspectos formales del caso frustren el acceso a la instancia superior. El alcance del recurso de casación fue luego extendido nuevamente con los fallos "Di Nunzio" (2005)[60], "Casal" (2005) y "Duarte" (2014)[61], por los cuales —muy resumidamente— la Corte determinó (i) que la CFCP debe intervenir en calidad de "tribunal intermedio" en toda causa que pueda eventualmente llegar a conocimiento de la CSJN; (ii) que la revisión debe realizarse en forma *amplia*, bajo el principio de máximo rendimiento, con la inmediación propia del juicio oral como único límite; y (iii) que la CFCP debe entender también en la revisión de condenas dictadas por ella misma, cuando implican la revocación de una absolución decidida en primera instancia.

En vista de aquellos hitos jurisprudenciales, uno esperaría quizás que el tópico analizado ("Alcance de la revisión") se disparara luego de cada fallo de la CSJN que propició un cambio importante en las atribuciones del tribunal de casación. Sin

[59] Art. 8.2 "h" de la Convención Americana sobre Derechos Humanos.
[60] Fallos: 328:1108.
[61] Fallos: 337:901.

embargo, lo que se ve es que el tema comienza a tomar impulso en los años *previos* al dictado de cada *leading case* de la CSJN.

Si bien la cuestión merece ser investigada con mayor profundidad, una explicación posible del comportamiento de la curva es que ella está describiendo los debates que ponen sobre la mesa las y los litigantes *antes* de que se expida la Corte. En efecto, si uno revisa los fallos en los que el tópico estudiado tiene mayor presencia, se encuentra con decisiones de la CFCP resueltas en sentido diametralmente opuesto al que luego adoptará la CSJN. Lo que parece verse, así, es que el debate pierde presencia cuando existe un fallo dirimente de la Corte, que la CFCP invariablemente tiende a acatar.

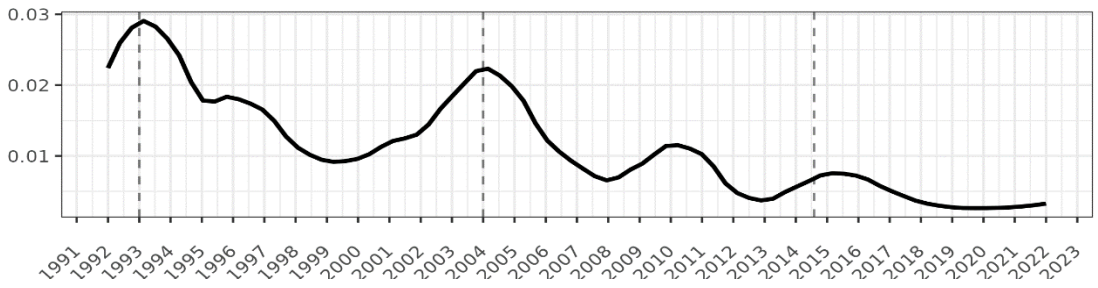


Figura 11: Tópico "Alcance de la revisión"

Por su parte, el comportamiento del tópico "Doctrina plenaria" a lo largo del tiempo parece decir algo sobre la relación que la CFCP ha tenido con ella misma; o, más precisamente, de la relación que han tenido las salas que la integran con las decisiones que el tribunal adopta en pleno y que, según el art. 10 de la ley 24.050, les son de aplicación obligatoria. La Figura 12 muestra en efecto tres momentos diferenciados en los que las salas hacen referencia a un fallo plenario —1992-1997, 1999-2002 y 2003-2009—, que se corresponden en gran medida con los períodos en los que la CFCP elaboró su doctrina plenaria. Lo más sobresaliente del gráfico a nuestro juicio, sin embargo, parece ser la observación de que luego del plenario n° 14 ("Blanc", de 2009), las referencias a los fallos plenarios decrecen notablemente. En efecto, en los últimos 14 años la CFCP no ha vuelto a dictar decisiones en pleno —una práctica muchas veces resistida por comentaristas y académicos, y que probablemente también se vio afectada por fallos de la CSJN como "Olariaga" (2007)[62] y "Acosta" (2008)[63], que de alguna manera desautorizaron algunas de las decisiones más controvertidas tomadas por el pleno de la casación federal—.

[62] Fallos: 330:2826.

[63] Fallos: 331:858.

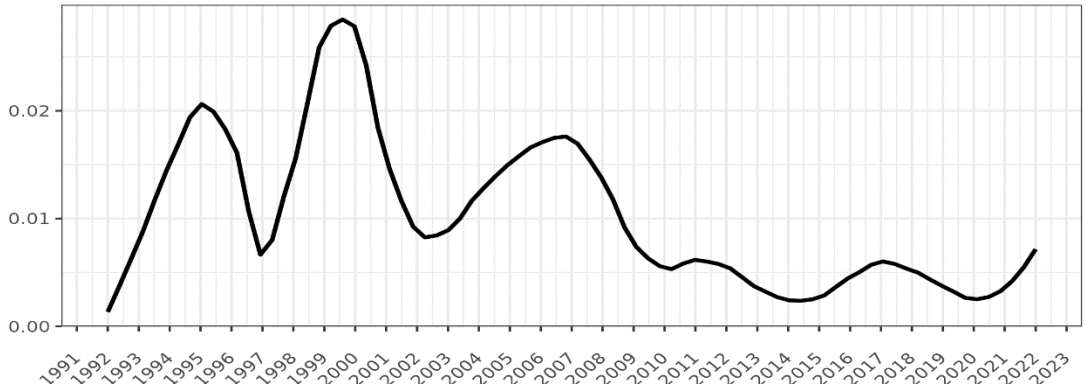


Figura 12: Referencias a la doctrina plenaria de la CFCP

4. Conclusión

Si bien la mayor parte de la discusión que actualmente vincula derecho e inteligencia artificial está enfocada en proveer los lineamientos jurídicos para la regulación de esta poderosa tecnología, en este trabajo intentamos mostrar que la relación entre ambas disciplinas puede ser también bidireccional; esto es, que las y los profesionales de la abogacía pueden amplificar sus capacidades de análisis jurídico adoptando la IA como herramienta tecnológica para estudiar y comprender el fenómeno, cada vez más complejo, en el que se está convirtiendo el derecho moderno.

Asimismo, darnos la oportunidad de aplicar algoritmos de IA para superar los desafíos metodológicos a los que la comunidad jurídica se enfrenta en la actualidad puede también ayudarnos a comprender con mayor profundidad cómo funcionan los sistemas inteligentes en la práctica y, de esa manera, permitirnos tomar parte del debate sobre su uso ético y centrado en las personas con un conocimiento más preciso y realista de sus riesgos y limitaciones.

Con este objetivo, en este capítulo hemos presentado una investigación exploratoria basada en la aplicación de un algoritmo de modelado de tópicos —*Latent Dirichlet Allocation* (LDA)—, que utilizamos para sintetizar, en pocas páginas, observaciones realizadas sobre una muestra de más de 80.000 sentencias pronunciadas por la Cámara Federal de Casación Penal a lo largo de sus 30 años de historia. Se trata de una tarea que, en la práctica, sería imposible de abordar mediante técnicas tradicionales —que exigirían, por ejemplo, anotar cada fallo no sólo de acuerdo a sus temas principales, sino en función de la presencia relativa de cada uno— . De hecho, el tamaño de la muestra analizada en este trabajo supera en varios órdenes de magnitud incluso a las estudiadas en investigaciones computacionales previas, realizadas sobre fallos de tribunales europeos y estadounidenses.

Esa circunstancia, sumada al hecho de que este es el primer trabajo que, de nuestro conocimiento, utilizó técnicas de este tipo para estudiar jurisprudencia penal en castellano, respaldan el potencial del *topic modelling* en general, y de LDA en particular, para dar cuenta de colecciones voluminosas de documentos en sí mismos

extensos y de alta densidad técnica —como lo son los fallos judiciales de un tribunal argentino—, muy alejadas de los casos de uso más populares, que típicamente analizan textos relativamente cortos (como los provenientes de redes sociales), y en idioma inglés.

Entre otras cosas, mediante el procesamiento computacional de la jurisprudencia de casación penal logramos observar, a título exploratorio, diversos aspectos del derecho penal argentino y de su máxima instancia especializada —la CFCP—, tales como los temas principales a los que se ha dedicado el tribunal a lo largo del tiempo, su relación con la Corte Suprema, o los delitos que más representación han tenido en sus decisiones, en diferentes momentos de su historia. También hemos podido documentar el abordaje judicial que han recibido ciertos temas especialmente relevantes, como aquellos vinculados con el juzgamiento de los crímenes contra la humanidad, la violencia machista o la trata de personas con fines de explotación.

En el camino, hemos abordado algunas confusiones conceptuales que suelen rodear la discusión sobre IA en ámbitos no especializados —como el jurídico—, delimitando con mayor precisión qué es (y qué no es) una inteligencia artificial, en su iteración más actual. Así, esperamos haber aportado herramientas y materiales para el desarrollo y florecimiento de la investigación jurídico-computacional, tanto en nuestro país como en el mundo, convencidos de que la IA está llamada a ocupar un lugar central en la abogacía del nuevo milenio.

5. Bibliografía

ALSCHNER, W., "The Computational Analysis of International Law", en DEPLANO, R. y TSAGOURIAS, N. (eds), *Research Methods in International Law: A Handbook* Edward Elgar Publishing, 2019.

BLEI, D.M., NG, A.Y., y JORDAN, M.I., "Latent Dirichlet Allocation", Journal of machine Learning research, vol. 3, 2003.

BRETT, M.R., "Topic Modeling: A Basic Introduction", Journal of Digital Humanities, vol. 2, 2012.

BURDICK, A., DRUCKER, J., LUNENFELD, P., PRESNER, T., y SCHNAPP, J., "Digital_Humanities", MIT Press, 2016.

CARTER, D., BROWN, J., y RAHMANI, A., "Reading the High Court at a Distance: Topic Modelling the Legal Subject Matter and Judicial Activity the High Court of Australia, 1903–2015", The University of New South Wales Law Journal, vol. 39, 2016.

CHURCH, K.W. Y HANKS, P., "Word Association Norms, Mutual Information, and Lexicography", Proceedings of the 27th Annual Meeting on Association for Computational Linguistics, 1989.

DOMINGOS, P., "A Few Useful Things to Know about Machine Learning", Communications of the ACM, vol. 55, 2012.

EISENBERG, T., "The Origins, Nature, and Promise of Empirical Legal Studies and a Response to Concerns", SSRN Journal, 2010.

GRIMMER, J. y STEWART, B.M., "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts", Political Analysis, vol. 21, 2013.

GRUS, J., "Data Science from Scratch", O'Reilly Media, 2015.

HOPP, C., "La lucha contra la trata de mujeres: la criminalización de las "malas víctimas" como consecuencia no deseada", en JARAMILLO, I. y CORREA, M.C. (eds), *Sexo, violencia y castigo* Didot, 2020.

JAMES, G.M., WITTEN, D., HASTIE, T., y TIBSHIRANI, R., "An Introduction to Statistical Learning", Springer, 2021.

KAPLAN, A. y HAENLEIN, M., "Siri, Siri, in My Hand: Who's the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence", Business Horizons, vol. 62, 2019.

KATZ, D.M., BOMMARITO, M.J., y BLACKMAN, J., "A General Approach for Predicting the Behavior of the Supreme Court of the United States", Plos One, vol. 12, 2017.

KOZLOWSKI, D., LOZANO RUBELLO, G., FELCHER, C.M., GONZALEZ, F., y ALTSZYLER LEMCOVICH, E.J., "Gender Bias in Magazines Oriented to Men and Women: A Computational Approach", ArXiv, 2020.

LIGHTHILL, J., "Artificial Intelligence: A General Survey", 1973.

MARTÍNEZ, E., "Re-Evaluating GPT-4's Bar Exam Performance", *SSRN*, 2023.

MEDVEDEVA, M., VOLS, M., y WIELING, M., "Using Machine Learning to Predict Decisions of the European Court of Human Rights", Artificial Intelligence and Law, vol. 28, 2020.

MIELNIK, D., "Análisis Computacional del Derecho Argentino", Revista Argentina de Teoría Jurídica, vol. 23, 2022.

MIELNIK, D., "Ciencia de Datos y Análisis Del Derecho Penal", Revista de Derecho Penal y Criminología, vol. 12, 2022.

NARAYANAN, A., BONNEAU, J., FELTEN, E., MILLER, A., y GOLDFEDER, S., "Bitcoin and Cryptocurrency Technologies: A Comprehensive Introduction", Princeton University Press, 2016.

NELSON, R.K., "Mining the Dispatch", University of Richmond, 2010.

PANAGIS, Y., CHRISTENSEN, M.L., y SADL, U., "On Top of Topics: Leveraging Topic Modeling to Study the Dynamic Case-Law of International Courts", Frontiers in Artificial Intelligence and Applications, vol. 294, 2016.

RÖDER, M., BOTH, A., y HINNEBURG, A., "Exploring the Space of Topic Coherence Measures", *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining* ACM, Shanghai China, 2015.

VARSAVA, N., "Computational Legal Studies, Digital Humanities, and Textual Analysis", *Computational Legal Studies* Edward Elgar Publishing, 2020.

VASWANI, A., SHAZEER, N., PARMAR, N., USZKOREIT, J., JONES, L., GOMEZ, I., "Attention is all you need", *ArXiv.Org*, 2017.

VELARDE, G., "Artificial Intelligence and Its Impact on the Fourth Industrial Revolution: A Review", ArXiv, 2020.