InDret
2.2026

Editorial

La inteligencia artificial generativa en la investigación criminológica cuantitativa

-

Uno de los grandes temas en la actualidad es, sin duda, el impacto del auge de la inteligencia artificial: tantas conversaciones cotidianas y profesionales giran sobre su gran capacidad para crear contenido audiovisual, las implicaciones para las actividades educativas o la posibilidad futura de un jefe supremo robot. Una cuestión relacionada que ha recibido poca atención en la criminología española (y el derecho español) es el uso de la inteligencia artificial generativa[12] en la investigación empírica cuantitativa. En el ámbito norteamericano, criminólogos, economistas y politólogos promueven un debate práctico, actualizado, iterativo y, por lo general, realista sobre los casos de uso de los nuevos modelos fundacionales.[13] A modo de ejemplo, véase, los blogs, Substacks y páginas web de Andrew Wheeler, Giovani Circo, Tyler Cowen y Alex Tabarrok, Scott Cunningham, Alexander Kustov o Andy Hall.[14] Con el fin de fomentar la conversación sobre el potencial y las limitaciones de las herramientas generativas para la investigación criminológica en España, contextualizaré la investigación cuantitativa con IA en nuestro ámbito y reflexionaré sobre algunas experiencias propias de utilizar modelos de lenguaje de gran tamaño (LLM) para la investigación.[15]

[12] La inteligencia artificial generativa se refiere a modelos de lenguaje de gran tamaño, como ChatGPT, Gemini, Claude, etc., que generan texto, imágenes y otros materiales (Narayanan & Kapoor, 2024). En el contexto de la investigación cuantitativa, se genera principalmente código para llevar a cabo análisis estadístico en lenguajes de programación como *R* o *Python*. Inteligencia artificial es un concepto genérico, ya que tiene cientos de miles de aplicaciones de diversa índole. La otra gran categoría de aplicaciones sería para predecir, por ejemplo, el clima en los próximos días, la duración de un viaje en Google Maps, el contenido que más se ajusta a nuestro perfil en redes sociales o la probabilidad de revictimización.

[13] Modelos fundacionales son modelos de IA entrenados con cantidades masivas de conjuntos de datos que se pueden utilizar como base para aplicaciones más específicas. Los LLM son ejemplos de modelos fundacionales.

[14] El blog de Andrew Wheeler está disponible en https://andrewpwheeler.com/. Giovani Circo se puede leer en https://gmcirco.github.io/blog/. El blog de Marginal Revolution es https://marginalrevolution.com/. Scott Cunningham tiene un *substack* y un podcast: https://www.scunning.com/. Alexander Kustov escribe en https://alexanderkustov.org/newsletter/. Los contenidos de Andy Hall se encuentran en su *substack*: https://substack.com/@andybhall

[15] Ninguna parte del presente texto ha sido generado o retocado con IA generativa, cosa que explica el estilo no nativo que no es lo más pulido. Pero para este formato prefiero mi 'voz' con sus idiosincrasias.

No existe mucha literatura sobre este tema porque los LLM son nuevos y el sistema de publicación académica no se caracteriza por su velocidad. Además, dada la rápida evolución de las capacidades de los LLM, es probable que los estudios metodológicos sobre ellos realizados en 2024 y publicados un año después (o más) se hayan quedado obsoletos durante el proceso de revisión y edición. Aun así, podemos encontrar algunos estudios europeos que sientan las bases para reflexionar sobre la intersección entre la investigación criminológica y la inteligencia artificial generativa. Gian Maria Campedelli publicó una revisión sistemática de estudios sobre inteligencia artificial e investigación sobre el crimen (Campedelli, 2021) y, en 2022, *Machine Learning for Criminology and Crime Research: At the Crossroads* (Campedelli, 2022), pero estas obras son anteriores al lanzamiento público de los modelos de lenguaje de gran tamaño como ChatGPT, Claude o Gemini. Sobre los LLM en concreto, Drápal y colegas han testado su aplicación al análisis temático (Drápal et al., 2023), pero de momento no han publicado sobre la investigación cuantitativa. A nivel español, Castro-Toledo et al. (2023) analizaron algunas cuestiones epistemológicas relacionadas con herramientas predictivas en el sistema penal, destacando la importancia de enfoques basados en la teoría para los algoritmos predictivos. El único estudio español que conozco que ha puesto el foco en los casos de uso de los nuevos modelos generativos es el de Sampayo Sande y Martínez Almanza, que hace una interesante clasificación automatizada de discursos parlamentarios para analizar las tendencias en la polarización del debate penal (2025). El método es innovador, pero, como bien señalan las autoras, no incluyó una validación formal que permitiría evaluarlo en comparación con humanos o los modelos de 2026.

Dos de los motivos por los que existe poca conversación sobre el uso de LLM en la criminología española son la calidad de los datos públicos y el nivel de la formación en métodos de investigación en general y en métodos cuantitativos específicamente. En primer lugar, respecto a la calidad de los datos, muchos de los ejemplos de uso en las fuentes norteamericanas señaladas en el primer párrafo utilizan la IA agéntica[16] para obtener y analizar datos públicos. Pero, ¿qué datos criminológicos oficiales se podrían utilizar en España? Como señalamos en un artículo reciente, ninguna institución española sabe ni de manera aproximada cuántas personas han sido condenadas por violación en España (Kemp et al., 2025). Si es así para un delito de tanta trascendencia social, poco nos podemos fiar de las cifras de los demás delitos. Otros autores también han destacado problemas de fiabilidad respecto a los datos sobre la actividad policial (López-Riba, 2021) o las dificultades para investigar con datos sobre el sistema penitenciario (Martí et al., 2021). No tiene mucho sentido automatizar la recopilación de datos agregados de poca calidad.

En segundo lugar, en términos de métodos cuantitativos, el nivel de formación de los egresados en Criminología es bastante bajo.[17] En mi universidad, ofrecemos más asignaturas metodológicas que muchas otras universidades, pero, debido a la amplitud de la criminología y los conocimientos previos del estudiantado, en lo cuantitativo tenemos solo una asignatura obligatoria en *Excel* y una optativa de programación básica en *R*. Aun así, me sorprendería si se

[16] Aunque no exista una definición consensuada, a grandes rasgos, la IA agéntica se refiere a los flujos de trabajo con IA y los agentes de IA. Los flujos de trabajo con IA son sistemas que utilizan LLM para ejecutar acciones predeterminadas para conseguir un objetivo. Los agentes de IA hacen lo mismo pero con mayor autonomía y razonamiento sobre los procesos y herramientas que se utilizan.

[17] Me imagino que el problema se extiende a diversos otros grados.

enseñan *R* o *Python* en más de tres de los otros cuarenta grados en Criminología en territorio español. Los estudios de másteres tampoco pueden ofrecer mucho más, ya que muchos estudiantes entran con formación escasa en métodos de investigación y, por tanto, se limitan a repasar lo que se ha hecho en los grados.

Dicho esto, los dos motivos que acabo de señalar se relacionan precisamente con dos áreas donde ciertas herramientas de IA generativa podrían ayudar (aunque con matices importantes que destacaré al final). Por un lado, si los investigadores en criminología suelen tener poca formación en lenguajes de programación para análisis estadístico, la IA generativa puede reducir las barreras de acceso. Hoy en día, los LLM acompañan a los programadores más experimentados de cualquier disciplina. Programar, en el sentido de dar instrucciones a aplicaciones informáticas, es difícil y lento, y los agentes de codificación pueden generar código mucho más rápido y con menor cantidad de los errores típicos que hacemos los humanos. Asimismo, los LLM pueden ayudar con la definición de la estrategia de análisis y las pruebas de robustez.[18] La criminología española tiene una tarea pendiente en lo que respecta a las pruebas de robustez y en la actualidad no hay excusa. Por otro lado, respecto al problema de los datos en España, uno de los casos de uso más claros de los modelos fundacionales es la capacidad para automatizar la extracción de datos de documentos no estructurados a gran escala y con mayor precisión que los humanos (Wheeler, 2026). El sistema penal español, como cualquier otro, documenta una diversidad de procesos y captura los resultados mediante, por ejemplo, atestados policiales, expedientes penitenciarios o resoluciones judiciales. Se puede, en principio, entrenar agentes de IA para extraer datos de cualquier forma de documentación y crear grandes conjuntos de datos para la investigación criminológica cuantitativa.

Resumidas las posibilidades para aplicar la inteligencia artificial generativa para ayudar con el análisis estadístico o con la creación de conjuntos de datos, es preciso introducir algunos matices, que lamentablemente son un poco circulares dado que tienen relación con los conocimientos y capacidades de los humanos. En el caso de apoyo para el diseño de estrategias de análisis o la programación, el problema reside en que es imprescindible tener cierto nivel de conocimientos sobre la investigación cuantitativa para elegir la estrategia de análisis óptima para responder a la pregunta de investigación en cuestión y para valorar los resultados. Si el LLM recomienda utilizar una regresión logística penalizada de Firth en lugar de un análisis de clases latentes o una regresión de Poisson, hay que saber valorar las diferentes opciones en función de la pregunta de investigación y los datos. Además, tampoco podemos dar por hecho que, aunque el código para el análisis sea perfecto, los resultados serán fiables. A modo de ejemplo, revisando recientemente los resultados de un análisis hecho con código generado parcialmente por un LLM, pude observar que el resultado de una variable era inverso a lo esperado. La mano de obra humana con conocimiento del contexto teórico era necesaria para identificar que en la limpieza de los datos las categorías de esta variable se habían invertido, dando lugar a un resultado incorrecto.

La labor humana también constituye la base de cualquier intento de automatizar la creación de conjuntos de datos con máxima precisión porque el proceso de entrenamiento es complejo, como hemos experimentado en el proyecto Desafíos empíricos de la Ley Orgánica de Garantía Integral

[18] Pruebas de robustez evalúan las asunciones adoptadas en una estrategia de análisis. Por ejemplo, se podría estimar un tipo de modelo distinto, incluir otra variable de control o configurar la variable dependiente con dos categorías en lugar de cuatro para ver si los resultados cambian.

de Libertad Sexual (LISEMPIRIC)[19] cuando hemos automatizado la extracción de datos de resoluciones judiciales a gran escala. Subir múltiples sentencias a un LLM y pedir ciertos datos necesita un largo proceso de entrenamiento para realizar la tarea de manera precisa: ¿el LLM tiene que sumar todas las condenas en casos con más de una pena o de un agresor? ¿Puede diferenciar atenuante A de atenuante B incluso cuando los magistrados hacen referencia a estas atenuantes con palabras que no son las mismas que el código penal? Si se quiere saber cuándo se denunciaron los hechos, ¿al LLM se le han dado instrucciones para no confundir el acto de denunciar formalmente con la presencia de la policía en el lugar de los hechos? ¿Sabe que las cabeceras de las resoluciones judiciales disponibles en algunas de las bases de datos más importantes están plagadas de errores de indexación que conllevan resultados erróneos en la extracción? Y así durante meses, concordando con el resumen de Andrew Wheeler (2026, p. 10):

> "This is a common story with many LLM tools. To make them work effectively for any particular application is not trivial"

En definitiva, la incorporación de herramientas de inteligencia artificial generativa para una investigación cuantitativa más robusta en la criminología española puede ser positiva siempre que vaya acompañada con un nivel de formación suficiente para evaluar los *outputs*. Asimismo, los modelos de lenguaje de gran tamaño permitirán crear conjuntos de datos muy poco comunes en nuestra comunidad, pero necesitarán un entrenamiento meticuloso por parte de personas especializadas en métodos de investigación social y en el contexto concreto del objeto de estudio en cuestión, es decir, una supervisión humana experimentada y detenida que sabe hacer los *prompts* idóneos e interpretar las respuestas. Si no hay esta supervisión humana adecuada, se aumenta la probabilidad de llegar a conclusiones poco fiables sobre temas de gran sensibilidad. No cabe duda de que el uso de LLM por parte de investigadores va a aumentar; por tanto, la criminología española debe reflexionar sobre cómo maximizar la calidad científica de los resultados.

Steven Kemp

[19] Financiado por la Agencia Estatal de Investigación. Ref: PID2023-152017NB-I00.

Bibliografía

Campedelli, Gian Maria. (2021). Where are we? Using Scopus to map the literature at the intersection between artificial intelligence and research on crime. *Journal of Computational Social Science*, 4(2), 503-530. https://doi.org/10.1007/s42001-020-00082-9

Campedelli, Gian Maria. (2022). *Machine Learning for Criminology and Crime Research: At the Crossroads*. Routledge, Taylor & Francis Group.

Castro-Toledo, Francisco J.; Miró-Llinares, Fernando y Aguerri, Jesús C. (2023). Data-Driven Criminal Justice in the age of algorithms: Epistemic challenges and practical implications. *Criminal Law Forum,* 34(3), 295-316. https://doi.org/10.1007/s10609-023-09454-y

Drápal, Jakub; Westermann, Hannes y Savelka, Jaromir. (2023). Using Large Language Models to Support Thematic Analysis in Empirical Legal Studies. En *Legal Knowledge and Information Systems* (Vol. 379, p. 197-206). IOS Press. https://doi.org/10.3233/FAIA230965

Kemp, Steven; Varona, Daniel y López-Riba, José Maria (2025). El misterioso caso de las cifras oficiales sobre condenas por delitos sexuales graves. *Revista Española de Investigación Criminológica,* 23(1), e1019-e1019. https://doi.org/10.46381/reic.v23i1.1019

López-Riba, José Maria (2021). El análisis cuantitativo de las identificaciones y las detenciones policiales en España: Datos disponibles, limitaciones e implicaciones: *Revista Española de Investigación Criminológica*, 19(2), Article 2. https://doi.org/10.46381/reic.v19i2.499

Martí, Marta; Güerri, Cristina y Pedrosa, Albert (2021). Fuentes de datos para la investigación criminológica en el ámbito penitenciario en España. *Revista Española de Investigación Criminológica,* 19(2), 1-31. https://doi.org/10.46381/reic.v19i2.515

Narayanan, Arvind y Kapoor, Sayash. (2024). *AI Snake Oil: What Artificial Intelligence Can Do, What It Can't, and How to Tell the Difference*.

Sampayo Sande, Sara y Martínez Almanza, Rocío. (2025). Polarización en el debate legislativo penal: Un análisis automatizado para el caso de la ley "solo sí es sí" en España. *Revista General de Derecho Penal,* (44), 13.

Wheeler, Andrew P. (2026). *Large Language Models for Mortals book: A Practical Guide for Analysts with Python.* Crime De-Coder. https://andrewpwheeler.com/2026/02/11/large-language-models-for-mortals-book/