



De la innovación a la ética: pautas de uso de la inteligencia artificial en la función legislativa

From Innovation to Ethics: Guidelines for the Use of Artificial Intelligence in the Legislative

Javier Guillem Carrau

Letrado de las Cortes Valencianas en servicios especiales

<https://orcid.org/0000-0002-4695-3582>

Fecha de recepción: 25/09/2025

Fecha de aceptación: 04/12/2025

Sumario: RESUMEN.—ABSTRACT.—I. INTRODUCCIÓN.—II. LOS MODELOS DE LENGUAJE MASIVOS Y LA FUNCIÓN LEGISLATIVA.—III. LOS DESAFÍOS ÉTICOS DEL USO DE LOS MODELOS EXTENSOS DEL LENGUAJE EN EL ÁMBITO PARLAMENTARIO.—3.1. El principio de igualdad.—3.2. El principio de veracidad.—3.3. El principio de privacidad.—3.4. El principio de responsabilidad.—3.5. El principio de legalidad.—IV. LA INCORPORACIÓN RESPONSABLE DE LOS MODELOS EXTENSOS DE LENGUAJE EN EL ÁMBITO PARLAMENTARIO.—4.1. La incorporación a las tramitaciones de textos legislativos y el rol de las Mesas como órgano de gobierno interior.—4.2. Las cajas blancas y la explicabilidad de los modelos.—4.3. La definición de la audiencia o público de los modelos extensos de lenguaje.—4.4. Las características y los escenarios alternativos.—V. CONCLUSIONES.—VI. BIBLIOGRAFÍA.

RESUMEN

Las oportunidades y riesgos que conlleva el uso de los modelos extensos de lenguaje en la generación de textos legales permiten identificar los mecanismos técnicos y las consideraciones estratégicas para su uso ético de acuerdo con los principios de igualdad, veracidad, privacidad, responsabilidad y legalidad. Se analizan los modelos de caja blanca, la explicabilidad de los modelos, la definición de la audiencia o público de estos y las conocidas como «características» y los «escenarios alternativos» y se propone una estrategia sobre el uso de la inteligencia artificial generativa en contextos de producción legislativa.

PALABRAS CLAVE: *Inteligencia artificial generativa, ética, legislación, Parlamentos y modelos de lenguaje.*

ABSTRACT

The use of LLM to generate legal texts presents both opportunities and risks. By examining white-box models, model explainability, target audiences, and alternative scenarios, we can identify the technical and strategic considerations necessary for their ethical use by respecting equal treatment, truthfulness, privacy, responsibility and legality. This analysis informs a proposed strategy for leveraging generative artificial intelligence in legislative contexts.

KEYWORDS: Artificial intelligence, ethics, legislation, parliaments, language models.

I. INTRODUCCIÓN

La inteligencia artificial generativa (IA generativa) emergió a finales del año 2022 como una nueva forma de crear contenidos y el uso de la misma a través de los llamados modelos extensos de lenguaje aplicados a la actividad humana promete una mejora de la productividad, de las tareas laborales y de la calidad del trabajo pero, a su vez, genera incertidumbres por sus implicaciones éticas y sociales.

Los Parlamentos han estado bregando con la tarea de afinar y adaptar la legislación sobre inteligencia artificial que busca un equilibrio entre innovación y supervisión responsable¹. Conforme la IA generativa va madurando, proliferan los casos confesos o inconfesos de uso en los Parlamentos mediante el empleo de los modelos extensos de lenguaje (conocidos como LLM, sus siglas en inglés), es decir, de programas informáticos conocidos como *ChatGpt*, *Geminí*, etc². En el marco de la actividad legislativa, el salto que el procedimiento legislativo puede experimentar por la aplicación de tecnologías o herramientas basadas en la inteligencia artificial generativa es de gran relevancia, aunque con efectos imprevisibles (Guillem Carrau, 2024, p. 78).

El análisis estratégico de la IA ha ocupado la atención hasta la fecha en el ámbito parlamentario³. En estas líneas se ahonda en el desafío que supone

¹ La Unión Interparlamentaria ha desarrollado una labor de documentación, bajo el concepto de «Parliamentary actions on AI policy», de todo tipo de iniciativas sobre IA que se desarrollan en los Parlamentos nacionales, incluyendo *revisões legislativas, informes, resoluções, perguntas, grupos de trabalho y discussões políticas*. A estos efectos, véase <https://www.ipu.org/impact/democracy-and-strong-parliaments/artificial-intelligence/parliamentary-actions-ai-policy>

² Al igual que la Unión Parlamentaria internacional, a los efectos de este artículo, circunscribimos el análisis a la inteligencia artificial generativa entendida como una forma de IA a la que se le puede pedir que genere texto, imágenes u otros datos por medio de modelos generativos basados en grandes volúmenes de datos (a menudo denominados «modelos extensos de lenguaje»). A estos efectos, véase <https://www.ipu.org/impact/democracy-and-strong-parliaments/artificial-intelligence/parliamentary-actions-ai-policy>

³ Al respecto véanse los recientes análisis realizados por: de Alba Bastarrechea, E. (2025). La inteligencia artificial en la transformación digital de los Parlamentos: en busca de un modelo ad hoc de go-

la inteligencia artificial generativa en la función legislativa, que se configura como un tema clave de futuro en la calidad de la legislación. Su uso en la generación de textos legales invita a identificar desafíos éticos, mecanismos técnicos y consideraciones estratégicas en el marco de un conjunto de principios jurídicos y preceptos legales que le son de aplicación, entre ellos el Reglamento europeo de inteligencia artificial⁴. Las conclusiones derivadas de dicha revisión permitirán fundamentar una hoja de ruta de la incorporación de la IA generativa al proceso legislativo.

II. LOS MODELOS DE LENGUAJE MASIVOS Y LA FUNCIÓN LEGISLATIVA

El salto que el procedimiento legislativo puede experimentar por la aplicación de tecnologías o herramientas basadas en la inteligencia artificial generativa es notable. En el desempeño de sus funciones constitucionales o estatutarias, los Parlamentos han sido hasta ahora libres de decidir cuando han querido confiar en la inteligencia artificial. A falta de una planificación estratégica, conforme avanza el uso de estas tecnologías, la libertad inicial que se predica de la autonomía parlamentaria quedará progresivamente reducida (Guillem Carrau, 2024, p. 77).

Una panorámica de la utilización de los modelos de lenguaje masivos en la función legislativa nos muestra que éstos pueden ser utilizados para una multitud infinita de tareas variopintas que abarcan desde analizar preceptos de textos legislativos e identificar patrones y relaciones entre diferentes leyes y regulaciones sobre la misma materia en distintos territorios, a predecir los resultados de diferentes escenarios legislativos. Además de apoyar la función legislativa con carácter global, estos modelos con supervisión humana se configuran como el instrumento para el análisis el impacto de las propuestas legislativas porque pueden analizar gran cantidad de referencias legislativas, tendencias e ideas y recomendaciones interesantes algoritmos (Fitsilis *et al.*, 2021, p. 27) y suponer, asimismo, una ayuda en la evaluación del impacto de la legislación; es decir, al ser capaces de archivar y analizar grandes volúmenes de legislación, pueden analizarlos y ayudar a identificar su impronta particular e identificar la justificación de la mayor o menor eficacia de la legislación.

beranza. *Asamblea. Revista Parlamentaria de la Asamblea de Madrid*, n.º 48, pp. 161–193; Guillem Carrau, J. (2024): Inteligencia artificial y Parlamentos modernos en *Corts: Anuario de derecho parlamentario*, ISSN 1136-3339, n.º 38, pp. 77-114.

⁴ Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial), «DOUE» n.º 1689, de 12 de julio de 2024.

Por ejemplo, una vez iniciado el proceso legislativo, los modelos de lenguaje masivos son de gran utilidad para el análisis de las enmiendas y de su correcta incorporación a los textos legales puesto que con las preguntas adecuadas (conocidas como *prompts* en sus siglas en inglés) se pueden extraer resultados muy útiles respecto a la correcta incardinación de las mismas en los preceptos legislativos (Guillem Carrau, 2024, p. 94).

Además, como inspirador del diputado o diputada que busca en ellos referencias cruzadas, pueden facilitar no solo opciones legislativas basadas en una síntesis de fuentes legislativas diversas sino ayudar a identificar los actores sociales relevantes y a mejorar y sacar más partido a la participación ciudadana en las fases de audiencia pública que garantizan los reglamentos parlamentarios en el procedimiento legislativo. Con este fin, pueden usar bases de datos para identificar a los expertos, actores sociales más relevantes, informes de investigación y otras fuentes relevantes de información.

No todo son ventajas como las enunciadas pues estas herramientas de inteligencia artificial generativa también tienen limitaciones, como la posibilidad de sesgo y discriminación en los datos de entrenamiento. Introduce nuevas complejidades y riesgos que deben ser cuidadosamente gestionados y evitados⁵.

Los modelos extensos de lenguaje no son un programa informático más como los que, hasta ahora, se caracterizaban por permitir el examen y la modificación de estos. Los citados modelos o códigos ya no son programas informáticos típicos o clásicos porque escapan del control humano y, básicamente, el trabajo previo relativo a la programación no dicta el resultado. En términos divulgativos, se explica que existe una caja negra que no nos cuenta lo que ha pasado por lo que no pueden ser auditados en el sentido clásico y no permiten hacer ingeniería inversa con los resultados obtenidos.

En el contexto de la inteligencia artificial generativa y de los modelos extensos de lenguaje, las decisiones fiduciarias pueden ser especialmente complejas debido a la capacidad de estos programas informáticos para aprender y adaptarse de manera autónoma. Esto plantea desafíos éticos y regulatorios que se resumen en torno a cómo se pueden diseñar modelos extensos de lenguaje que tomen decisiones fiduciarias de manera responsable y transparente y cómo se pueden establecer límites y controles para garantizar que los modelos extensos de lenguaje no tomen decisiones que perjudiquen a sus usuarios o a la sociedad en general. Estos interrogantes permiten subrayar la importancia de las características del modelo: la importancia de las características se refiere a la medida en que cada variable o característica de un conjunto de datos contribuye a la precisión o eficacia de un modelo de aprendizaje automático.

Entrando en el siguiente punto de este análisis, examinemos el perfil de riesgo de cada una de estas tareas: ¿qué salvaguardias y normas internas ten-

⁵ Inter-parliamentary Union (2024): Guides for AI in parliaments en <https://www.ipu.org/ai-guidelines>, p. 17.

drán que implementarse para que cualquiera de estas tareas se pueda confiar a la IA? El nivel de impacto influye: es probable que sean tareas relativamente bien definidas que los sistemas de IA puedan utilizar con poco o ningún riesgo. En otros casos, como el de la redacción de legislación, las posibles repercusiones podrían ser muy altas, lo cual tendría un gran impacto en la vida de las personas⁶.

Como se describe en estas líneas, entre los riesgos y debilidades de los modelos extensos de lenguaje es necesario citar la existencia de sesgos y riesgos de seguridad y de privacidad; las cajas negras o el déficit de explicabilidad y las alucinaciones y los costes elevados y su impacto medioambiental. Asimismo, es conocido que estos modelos generan una dependencia excesiva y ponen en juego una serie de principios éticos que subrayan la necesidad de hacer evolucionar las capacidades de sus señorías y del personal al servicio en las cámaras legislativas.

En lo que concierne a los sistemas de inteligencia artificial, incluidos los modelos extensos de lenguaje, como veremos posteriormente, el Reglamento europeo de inteligencia artificial anteriormente citado, establece un esquema de reconocimiento de los riesgos asociados a la utilización de la inteligencia artificial y describe unas categorías en función de las cuáles procede un tipo u otro de actuaciones. Esta clasificación es de gran utilidad en cualquier planteamiento organizativo relativo a la IA pero hay que añadir que, a escala internacional, es ya de referencia una norma ISO para manejar los riesgos asociados a la inteligencia artificial⁷.

III. LOS DESAFÍOS ÉTICOS DEL USO DE LOS MODELOS EXTENSOS DE LENGUAJE EN EL ÁMBITO PARLAMENTARIO

Como cualquier otro campo de la experiencia humana, la inteligencia artificial en cualquiera de sus desarrollos no está exento de debates éticos y de riesgos asociados a la posibilidad de falta de objetividad, la parcialidad del análisis basado en la información de la que se alimentan, los sesgos o la discriminación no intencionada y los incidentes derivados de los mismos (Guillem Carrau 2024, p. 105).

Por ejemplo, las llamadas «alucinaciones» algorítmicas que suponen la creación de contenido falso inadecuado o irrelevante por la inteligencia artificial en su versión generativa supone un riesgo que llama a la necesidad de aplicar pensamiento humano crítico y buenas prácticas consolidadas (Dahl y otros, 2023).

⁶ Véase *op. cit.* 5.

⁷ Véase <https://www.iso.org/standard/77304.html>

Sobre la base de este planteamiento se aborda a continuación el desafío ético de los modelos extensos de lenguaje, como cualquier otra herramienta de inteligencia artificial generativa puesta al servicio de la actividad parlamentaria, que debe cumplir una serie de principios éticos y legales con una perspectiva amplia que abarque no sólo su uso ético y responsable sino también inspiren su diseño para que la creación y desarrollo de estos tenga en cuenta los principios de igualdad, transparencia, privacidad, veracidad, responsabilidad y legalidad.

Se trata de unas garantías mínimas que no pueden quedar al margen de la incorporación de las nuevas tecnologías de la inteligencia artificial a los Parlamentos (Guillem Carrau, 2024, p. 104). En todo caso, la libertad de elección respecto al uso de cualquier forma de inteligencia artificial en los Parlamentos no está exenta de responsabilidades y deberes por las implicaciones legales y éticas que conlleva (Montoya y Rummery, 2020, p. 10). Además, su implementación sin el debido control jurídico puede derivar en riesgos graves para los derechos fundamentales y para la legitimidad de las instituciones democráticas (de Alba, 2025, p. 167).

Un ejemplo ya citado en anteriores publicaciones (Guillem Carrau, 2024, p. 107) del debate ético al respecto de una herramienta parlamentaria basada en la inteligencia artificial es el proyecto del Parlamento flamenco que usó la inteligencia artificial para detectar los miembros del Parlamento que no están atentos al debate al usar sus teléfonos móviles en las sesiones plenarias⁸. Al respecto, es preciso identificar cuáles son los principios que inspiran este tipo de sistemas de control, si los sistemas a poner en práctica son proporcionales al fin perseguido y si pueden redundar en una mejor consecución de los principios que los inspiran. En concreto, es necesario evaluar si esta clase de mecanismos de control pueden ser un mecanismo adecuado para fomentar la participación en los debates parlamentarios o no.

En definitiva, los modelos extensos de lenguaje generan un conjunto de desafíos éticos por el mero hecho de ser una tecnología, en primer lugar; por el hecho de ser nueva, en segundo lugar; y por el hecho de que el empleo de la misma produce resultados no controlados por su creador, en tercer lugar.

Por ello, como se analiza a continuación, se deben asociar los principios de igualdad, veracidad, seguridad, responsabilidad y legalidad a cualquier planteamiento frente a los desafíos éticos que nos plantea el uso de estas herramientas en los entornos parlamentarios.

⁸ Véase <https://aibusiness.com/verticals/this-ai-bot-calls-out-distracted-belgian-politicians-who-spend-too-much-time-on-their-phones>, accesible el 21.07.2025.

3.1. El principio de igualdad

Como con cualquier otra tecnología, la inteligencia artificial generativa genera un interrogante ético en torno al respeto del principio de igualdad en dos campos, el de acceso a la tecnología y el de la posibilidad de que el uso de la misma perpetúe desigualdades o acreciente los problemas generados por las mismas.

Así, en primer lugar, respecto al problema de acceso a las tecnologías, la razón de este cuestionamiento se encuentra en la posibilidad de acceso a la misma y, consecuentemente, de la posibilidad de obtener los beneficios que el empleo de esta conlleva en sus tareas personales o profesionales.

Es evidente que aquellos sujetos que sepan manejarla y emplearla a su favor va a obtener una ventaja competitiva lo cual plantea el debate ético del acceso a la misma. En dicho contexto, los sujetos con capacidad de interactuar y gestionar a su interés las herramientas de inteligencia artificial generativa tendrán eventualmente mejores resultados personales y profesionales que aquellos sujetos que no lo tengan.

Trasladado este debate al ámbito parlamentario, respecto a este derecho a la igualdad en materia de educación en la era digital, las administraciones parlamentarias tienen la responsabilidad de formar tanto a los miembros de las asambleas legislativas y sus asesores en los grupos parlamentarios, como a los propios funcionarios de las instituciones en el conocimiento, gestión y aprovechamiento de estas nuevas tecnologías de manera aplicada a las tareas que tienen encomendadas en función de su mandato parlamentario, de las relaciones contractuales o de las funciones que tienen asignadas en las respectivas plantillas o relaciones de puestos de trabajo.

Este esfuerzo organizativo de formación permanente se advierte como elemento clave para que se respete el desafío ético que supone el respeto al principio de igualdad en el contexto de una administración parlamentaria. Una apuesta decidida en este sentido significará que todos estos sujetos serán capaces de fortalecerse gracias a la inteligencia artificial generativa y no quedarán excluidos de los avances o ventajas que su uso lleva aparejado.

En cuanto al contenido de estas formaciones y capacitaciones, cabe manifestar que deberán estar adaptadas a cada función o tarea pues no es lo mismo la familiaridad que se le tiene que exigir a un letrado o a los oficiales de gestión parlamentaria en el manejo de un modelo extenso del lenguaje para la revisión de enmiendas legislativas y su correcta incardinación en el texto legislativo que el que necesitan los ordenanzas, los técnicos de registro y sonido en el desarrollo de sus funciones de atención o de grabación de una sesión parlamentaria o los técnicos de mantenimiento que se ocupan de la gestión de la temperatura de las distintas estancias parlamentarias. Todos ellos producen o recogen datos en el desarrollo de sus funciones que son susceptibles de ser analizados y que pueden resultar claves, con ayuda de herramientas de inteligencia artificial generativa en la mejora de las funciones que desempeñan.

En segundo lugar, el desafío ético se plantea en relación con la perpetuación de las desigualdades por la utilización de la inteligencia artificial generativa y la posibilidad de que se convierta en factor de aceleración de estas o de perpetuación de las diferencias. Los sesgos asociados al uso de los modelos extensos de lenguaje vuelven a tener aquí un rol protagonista y la administración parlamentaria debe plantearse el uso de modelos extensos de lenguaje que funcionen como factor de desaceleración.

Es evidente que se deberá fomentar y publicitar el uso de modelos extensos de lenguaje que no favorezcan la obtención de resultados o soluciones sexistas o discriminatorias en razón de edad, sexo, raza, origen u otros motivos de discriminación frente a aquellos que no lo hagan. Asimismo, se revela fundamental que la utilización de estos modelos en los entornos de relación con la ciudadanía y el Parlamento sea respetuosa con el principio de igualdad e incluso contenga alguna medida de discriminación positiva para, por un lado, equilibrar los posibles sesgos que contengan, y por otro, facilitar acceso a los mismos para personas con discapacidades diversas.

3.2. El principio de veracidad

Señala Coeckelbergh (2021, p. 89) que la inteligencia artificial generativa puede conducir a nuevas formas de manipulación, vigilancia, y totalitarismo no necesariamente bajo la apariencia de régimenes autoritarios sino de manera subrepticia y altamente efectiva. Este desafío ético directamente relacionado con el principio de veracidad tiene un exponente claro en los debates políticos y en la creación de noticias falsas y bulos y el impacto que ello está teniendo en nuestras sociedades.

De este desafío no escapa el parlamentarismo tampoco y no hay que circunscribirse solo a los debates parlamentarios, la labor de los responsables de prensa de los grupos parlamentarios o del propio Parlamento y el incremento de la tensión en dichos entornos derivado el mundo interconectado en el que se desarrollan las sesiones parlamentarias, sino también en la puesta en duda y en la creación de incertidumbres sobre los elementos básicos de funcionamiento de las asambleas parlamentarios, por ejemplo, el mandato parlamentario frente a la función clave de los partidos políticos y los grupos parlamentarios, la función legislativa frente a la producción de normas vía decreto por los ejecutivos o la sustitución del pleno o las comisiones por otros foros como centro del debate político.

Es posible gestionar y mitigar los sesgos, aunque esto exige transparencia en el mismo centro de los programas informáticos en cuestión. La «explicabilidad» de la que se ha hablado anteriormente —esto es, la capacidad de entender cómo y por qué producen los sistemas de IA generativa determinados resultados— es vital porque garantiza la transparencia y la rendición de cuentas, y permite así una adopción de decisiones fundamentada y un uso ético. Los Parla-

mentos deberán demostrar que estos principios están integrados en su uso de la IA generativa para generar confianza entre sus miembros y el público. Probablemente, la legislación y los reglamentos tendrán dificultades para mantener el ritmo vertiginoso del desarrollo tecnológico. Sin embargo, en muchos casos, las leyes existentes sobre desinformación, campañas electorales, propiedad intelectual y difamación, por ejemplo, pueden servir de protección contra el uso indebido del contenido generado con los modelos extensos de lenguaje⁹.

Por ello, el respeto al principio de veracidad en el ámbito parlamentario está directamente asociado a los deberes de transparencia y accesibilidad. Se convierte en pieza clave del sistema la posibilidad de explicar cómo funciona el modelo. Esto ocurre por la sencilla razón de que deberá poder explicarse el sesgo que tiene el software y también advertir a los usuarios de este sesgo para que puedan ser conscientes, mitigar o asumir dicho sesgo, presente en los resultados y soluciones que ofrecerá el modelo extenso de lenguaje en cuestión.

En este sentido, cabe referirse a las directrices de inteligencia artificial de la Unión Interparlamentaria que contienen un conjunto de principios éticos que deben ser salvaguardados cuando se empleen estas técnicas en el ejercicio de la función legislativa. De este modo, son principios éticos a respetar los siguientes: la privacidad y la protección de datos personales; la transparencia y la dación de cuentas de modo accesible y controlable y alineado con los principios democráticos; la garantía de lealtad y no discriminación en la formación del personal y la gestión de datos, así como en la colaboración; la seguridad y solidez y respetar los principios de bienestar social y medioambiental; y por último, la autonomía y la libertad humana como medios de supervisión¹⁰.

A modo de ejemplo, se puede citar que las citadas directrices contienen una serie de ejemplos de acciones para salvaguardar los referidos principios éticos; así, para que los principios de lealtad y no discriminación sean respetados por los modelos extensos de lenguaje en la actividad legislativa se propone que se asegure que los datos no tienen sesgo ideológico y que no prejuzgan; que se sea consciente de los sesgos históricos cuando los datos provienen de la actividad de plenos y comisiones; que se coopere con entidades externas para obtener datos que ayuden a mantener la calidad del sistema; que se vigilen los sesgos derivados de traducciones; que se confirme que la información producida por GenAI no tiene sesgos antes de utilizarla y que se identifiquen los sesgos existentes en los datos provenientes de la participación ciudadana en el proceso legislativo. Asimismo, será necesaria la creación de un registro de los modelos extensos de lenguaje y otros programas de

⁹ Inter-parliamentary Union (2024): Guides for AI in parliaments en <https://www.ipu.org/ai-guidelines>

¹⁰ Inter-parliamentary Union (2024): Guides for AI in parliaments en <https://www.ipu.org/ai-guidelines>, p. 76 y ss.

ordenador o aplicaciones empleadas en el Parlamento y que desarrollen IA generativa. En dicho registro, se puede aprovechar para indicar todos estos aspectos relativos a la explicabilidad para respetar el principio de veracidad y transparencia.

3.3. El principio de privacidad

Siguiendo a Coeckelbergh (2021, p. 89) otro desafío ético de la inteligencia artificial es la propia privacidad y seguridad. Por la propia realidad del uso de los modelos extensos de lenguaje, existe un margen amplio de riesgo en que los dispositivos electrónicos o el propio programa informático sean hackeados o manipulados.

De Alba Bastarrechea (2025, p. 168) describe el riesgo específico en el ámbito parlamentario derivado de la recolección masiva de datos personales y su tratamiento con la inteligencia artificial generativa que plantea, en particular, injerencias ilegítimas en la privacidad, especialmente si se utilizan técnicas de elaboración de perfiles o decisiones automatizadas sin base jurídica suficiente. Asimismo, considera que la utilización de IA en la gestión documental plantea interrogantes sobre la autenticidad, integridad y conservación a largo plazo de los registros parlamentarios, esenciales para la preservación de la memoria democrática.

El desafío del principio de privacidad no tiene solo una faceta técnica de seguridad como la descrita sino que también genera un campo de acción en torno los tratamientos de datos personales en el entrenamiento, validación, despliegue, explotación y retirada de cualquier solución de modelos extensos de lenguaje. En este contexto, la protección de las personas físicas en relación con el tratamiento de datos personales es un derecho fundamental. El artículo 8, apartado 1, de la Carta de los Derechos Fundamentales de la Unión Europea y el artículo 16, apartado 1, del Tratado de Funcionamiento de la Unión Europea establecen que toda persona tiene derecho a la protección de los datos de carácter personal que le conciernen¹¹. El derecho fundamental a la protección de datos está desarrollado en un marco normativo que actualmente comprende el Reglamento europeo y la Ley Orgánica de protección de datos¹².

Como señala la Agencia Española de protección de datos los modelos extensos de lenguaje pueden manejar datos de personas físicas, como en un

¹¹ Considerando 1 del RGPD.

¹² Reglamento (CE) n.º 619/2016 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (RGPD), y se complementa en la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales (LOPDGDD) además de toda la normativa sectorial publicada antes y después de la entrada en vigor del RGPD.

modelo de perfilado de marketing o electoral, o puede haber tratamientos en los que no aparezcan datos de carácter personal, como podría suceder en un modelo de predicción meteorológico que recoge datos de estaciones geográficamente distribuidas.

En los desarrollos de modelos extensos de lenguaje aplicados a la actividad parlamentaria, siguiendo el documento citado, se deben respetar una serie de principios básicos que contiene la legislación en materia de protección de datos (licitud, lealtad y transparencia; limitación de la finalidad (especificación del propósito); minimización de datos; exactitud; limitación del plazo de conservación de las conversaciones con el modelo; la integridad y la confidencialidad. El respeto a los mismos permitirá afirmar que se dan las condiciones para un tratamiento legítimo de los mismos. Asimismo, se deberá garantizar el ejercicio de los derechos que asisten a los sujetos de los datos e implementar mecanismos efectivos para el ejercicio de éstos, en particular los derechos de transparencia, información, acceso, rectificación, supresión, limitación, oposición, portabilidad y, uno más que tiene gran importancia en determinadas aplicaciones de IA como son los derechos que tienen los ciudadanos con relación a la toma de decisiones automatizadas¹³.

Entre todos ellos, merece la pena referirse al derecho de información sobre cómo funciona el modelo extenso de lenguaje. En este sentido, la Agencia (2020, p. 24), aporta un ejemplo de información que podría tener relevancia de cara al interesado, sería: el detalle de los datos empleados para la toma de decisión, más allá de la categoría, y en particular información sobre los plazos de uso de los datos (su antigüedad); la importancia relativa que cada uno de ellos tiene en la toma de decisión; la calidad de los datos de entrenamiento y el tipo de patrones utilizados; los perfilados realizados y sus implicaciones; los valores de precisión o error según la métrica adecuada para medir la bondad de la inferencia; la existencia o no de supervisión humana cualificada; la referencia a auditorías, especialmente sobre las posibles desviaciones de los resultados de las inferencias, así como la certificación o certificaciones realizadas sobre el sistema de IA. En el caso de sistemas adaptativos o evolutivos, la última auditoría realizada; en el caso de que el sistema IA contenga información de terceros identificables, la prohibición de tratar esa información sin legitimación y de las consecuencias de realizarlo.

Finalmente, como en cualquier otro contexto de protección de datos, la Agencia (2020, p. 16) sugiere que la referencia debe ser un modelo responsabilidad y cumplimiento establecido basado en la «accountability», o responsabilidad proactiva, y cuyos elementos rectores son: la identificación de una responsabilidad en el tratamiento; el análisis del riesgo para los derechos y libertades; el estudio de la necesidad y la proporcionalidad de las operaciones de tratamiento con respecto a su finalidad; y el despliegue de medidas para la

¹³ AEPD (2020): Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial. Una introducción, disponible en <https://www.aepd.es/guias/adecuacion-rgpd-ia.pdf>

gestión del riesgo, medidas de privacidad por defecto y desde diseño, medidas de seguridad, de gestión de incidentes, etc.

En el ámbito parlamentario, este planteamiento frente al desafío ético de la privacidad genera la necesidad de la identificación de las distintas responsabilidades asociadas al tratamiento de los datos en el marco de la Administración parlamentaria, al igual que se hace con cualquier otro tratamiento de datos derivado de otras actividades del Parlamento. La base jurídica que legitimará dichos tratamientos será en algunos casos que el tratamiento es necesario para la ejecución del contrato informático en cuestión, en otros se alegará el interés legítimo¹⁴ o se precisará el consentimiento de los interesados¹⁵ aunque también pueden ser bases jurídicas las razones de interés público o el ejercicio de poderes públicos o el cumplimiento de obligaciones legales.

Aquí entran en acción las secretarías generales de las Asambleas parlamentarias y, en concreto, sus servicios informáticos y el delegado de protección de datos que deben garantizar la compatibilidad del uso de este tipo de herramientas con las medidas de seguridad implementadas en la arquitectura informática y las aplicaciones empleadas en sede parlamentaria, incluidos cualquier tipo de herramienta de las llamadas API (*Application Programming Interface*) que permite que diferentes sistemas y softwares de comuniquen e intercambien información de manera eficiente con unos estándares y criterios predefinidos; así como de la creación de los registros correspondientes y de su comunicación a la Agencia.

3.4. El principio de responsabilidad

El principio de responsabilidad genera un desafío de doble cara en el ámbito parlamentario al afectar tanto a la responsabilidad de sus creadores sobre el modelo extenso de lenguaje y sobre las creaciones de este. Este doble enfoque de la cuestión de la responsabilidad sobre y por la IA generativa invita a realizar una reflexión general sobre la propia autoría de los resultados de la utilización de los modelos extensos de lenguaje que escapa al objeto de estas líneas.

¹⁴ Véase el Dictamen 06/2014 sobre el concepto de interés legítimo, del Grupo del Artículo 29, desarrolló en detalle cómo evaluar los factores que legitiman el interés del responsable para realizar un tratamiento de datos personales y equilibrarlo con los derechos y los intereses, también legítimos, de los interesados accesible en https://www.aepd.es/media/criterios/wp217_es_interes_legitimo.pdf

¹⁵ En particular, el artículo 22.4 RGPD establece que las decisiones basadas únicamente en el tratamiento automatizado, incluida la elaboración de perfiles, que produzca efectos jurídicos en el interesado o le afecte significativamente de modo similar, no se basarán en las categorías especiales de datos personales salvo que medie el consentimiento del interesado o el tratamiento sea necesario por razones de un interés público esencial, sobre la base del Derecho de la Unión o de los Estados miembros y se hayan tomado medidas adecuadas para salvaguardar los derechos y libertades y los intereses legítimos del interesado.

No obstante, conforme a lo dicho, en primer lugar, el debate gira en torno a la responsabilidad sobre los modelos de inteligencia artificial generativa de sus creadores. Así, el principio de responsabilidad se manifiesta en la llamada a la responsabilidad sobre las creaciones de los científicos, la remarca Coeckelbergh (2021, pp. 28 a 30) al citar la novela de Frankenstein (Mary Shelley) y explicar el paralelismo existente entre la inteligencia artificial como desarrollo científico y el mensaje principal de la novela que consiste en que los científicos necesitan asumir la responsabilidad de sus creaciones. Por un lado, la novela subraya el claramente el peligro de la tecnología de la que se pierde el control y, en particular de los humanoides artificiales enloquecidos; y por otro, tiene una narrativa que incorpora la competición entre la creación artificial y la humana¹⁶.

De igual modo, en segundo lugar, el desafío se cristaliza en una segunda diatriba en torno a la posibilidad de considerar algún tipo de personalidad y por la tanto de responsabilidad a los propios modelos extensos de lenguaje y, en general, a cualquier software de inteligencia artificial generativa. Desde un punto de vista práctico, el otorgamiento de una mayor capacidad de actuación y la asunción de tareas que solían llevar a cabo los humanos por un sistema de inteligencia artificial parece llevar a la conclusión de que dichos sistemas, o sea cualquier modelo extenso de lenguaje, tiene atribuida cierta capacidad y, por lo tanto, responsabilidad. La respuesta práctica a este desafío estará directamente ligada a la transparencia del modelo y la explicabilidad de sus decisiones, aspectos que ya han sido analizados anteriormente en este artículo en relación a la posibilidad y a la exigencia moral de entender cómo ha decidido un modelo de lenguaje.

Por ello, en los Parlamentos, la respuesta al desafío ético de la responsabilidad de las producciones de los modelos extensos de lenguaje debe estar claramente relacionada con el desarrollo de mecanismos de cajas blancas para contrarrestar el problema de las llamadas cajas negras. Esto reforzará la confianza en su uso por los actores parlamentarios y la ciudadanía en general así como la responsabilidad de sus creadores sobre los citados modelos. La institución parlamentaria deberá ser capaz de responder a la pregunta de por qué el modelo ha tomado una decisión y no otra. En el uso ordinario en los ámbitos parlamentarios, una respuesta adecuada a este desafío permitirá explicar la razón de determinadas enmiendas en un proceso legislativo o el motivo de las preguntas formuladas por los grupos parlamentarios de la oposición al gobierno en la sesión de control.

Asimismo, la ciudadanía y los miembros del Parlamento deben tener la seguridad de que las normas son respetadas por los modelos extensos de len-

¹⁶ El llamado complejo de «Frankenstein» se pone de manifiesto en gran número de producciones culturales como las películas «2001. Odisea en el espacio» (1968) o la saga de «Terminator». Incluso científicos como Stephen Hawking han expuesto en público este complejo con declaraciones como esta: «El éxito en la creación de la inteligencia artificial podrá ser el evento más grande en la historia de la humanidad. Desafortunadamente también sería el último, a menos de que aprendamos cómo evitar los riesgos» (Hawking, 2014).

guaje y que existe una supervisión humana en su diseño y en el seguimiento de los resultados aportados por su utilización en las distintas funciones parlamentarias. Esto pasa por tener los modelos bajo supervisión humana y reducir los riesgos y sesgos a un nivel aceptable para la ciudadanía. Como se ha mencionado anteriormente, es clave mitigar la caja negra o hacerla blanca con la vigilancia humana (Korr, 2024) pero es fundamental para que se asegure el funcionamiento correcto de la aplicación de inteligencia artificial y evitar los fallos, por ejemplo, verificando la calidad de las explicaciones cuando sea necesario como en casos donde haya intereses económicos o jurídicos en juego. Respecto a las cajas negras, el elemento de la transparencia será otro requisito fundamental no solo respecto al trabajo interno de la aplicación sino respecto al reconocimiento del proceso de automatización, los datos de entrenamiento de la aplicación y los modos de cuestionar y desafiar las decisiones tomadas automáticamente (Miller, 2019).

De igual modo, será preciso que se cree un registro de modelos extensos de lenguaje para dar visibilidad a los reguladores en cómo estas tecnologías se desarrollan y se ponen en práctica. En este contexto, se hace necesaria la determinación previa de unos requisitos de verificación formal con protocolos y sistemas y además de verificación de su respeto al sentido común y la seguridad. En cualquier caso, se debe afirmar que no cabe la autorregulación por los propios modelos. Por eso, el uso de los modelos extensos de lenguaje en la actividad legislativa exige que se decida un tipo de supervisión humana (HITL, HOTL, HIC)¹⁷ y el diseño de una estrategia de desarrollo basada en valoración del riesgo, establecimiento de normas, creación de estándares y realización de tests, entrenamiento del personal, informes periódicos y revisión de resultados y procesos; creación de un canal digital para comentarios de usuarios del sistema; invertir en formación del personal; llevar a cabo comprobaciones humanas constantemente; y seleccionar cuidadosamente las aplicaciones a poner en marcha, sus características y que todos los usuarios comprendan sus limitaciones.

Si se toma con seriedad, el entrenamiento de estos sistemas de inteligencia artificial generativa guiado por humanos puede aliviar algunos de los problemas técnicos y éticos, mejorando enormemente con la intuición humana así como satisfacer la necesidad de explicarlos de manera más efectiva y equilibrada (Coglianese *et al.*, 2024).

De este modo, deberá ser posible intervenir en los sistemas de inteligencia artificial cuando proceda como parte del proceso de licencia de los programas informáticos y determinar mecanismos de autoregistro para que

¹⁷ Los tipos de supervisión humana de genAI se refieren a: HITL (Human-In-The-Loop), donde los humanos intervienen en el proceso de decisión de la inteligencia artificial; HOTL (Human-On-The-Loop), donde los humanos supervisan y corrigen el resultado de la inteligencia artificial después de que se ha tomado la decisión; y HIC (Human-In-Command), donde los humanos tienen el control total sobre la inteligencia artificial y toman decisiones finales basadas en la información proporcionada por la IA.

el regulador conozca de su existencia y pueda comprobar que se respetan las líneas rojas para los modelos extensos de lenguaje de acuerdo con el Reglamento europeo de IA.

Por último, en el uso diario por los miembros de las Asambleas se impone la necesidad de que se adopten prácticas responsables para citar cuando sus textos e intervenciones, incluso el texto de las enmiendas o proyectos legislativos han sido elaborados con la ayuda de estos sistemas generativos, incluso con la mención de que «prompt» se ha empleado para obtener el resultado deseado, es decir, que pregunta se ha planteado al sistema para obtener la respuesta empleada por el miembro de la asamblea legislativa. En el desarrollo de estas prácticas responsables o directrices de uso, será fundamental potenciar y garantizar la participación no solo de los actores parlamentarios sino de la ciudadanía representada por estos de manera que dichas directrices o prácticas tomen en consideración las aportaciones de todas las partes interesadas.

3.5. El principio de legalidad

Con la adopción del Reglamento (UE) 2024/1689, la UE puso de manifiesto un liderazgo mundial en torno a la adopción de normas para el uso de la inteligencia artificial generativa y lo ha intentado sostener con otros esfuerzos de gobernanza internacional como el proceso de Hiroshima del G7¹⁸, la declaración de Bletchley¹⁹, la declaración de París²⁰ y el Convenio marco del Consejo de Europa sobre inteligencia artificial²¹.

Precisamente, la aparición en los mercados del ChatGPT y equivalentes tuvo como consecuencia que el Parlamento Europeo tuviera que modificar el texto legal en tramitación parlamentaria para manifestar su punto de vista sobre estas nuevas herramientas; esto es la Comisión ya había redactado el Libro Blanco²² y el Parlamento Europeo se había manifestado sobre la regulación de la inteligencia artificial²³. En consecuencia, en el Reglamento (UE)

¹⁸ Véase <https://digital-strategy.ec.europa.eu/en/library/hiroshima-process-international-code-conduct-advanced-ai-systems>

¹⁹ Véase <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration-the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023>

²⁰ Véase <https://www.elysee.fr/en/manuel-macron/2025/02/11/statement-on-inclusive-and-sustainable-artificial-intelligence-for-people-and-the-planet>

²¹ Véase <https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence>

²² En 2020, el Libro Blanco sobre la inteligencia artificial supuso el primer enfoque de la Comisión sobre la materia. Este documento definió las opciones existentes para alcanzar el doble objetivo de promover la adopción de la IA y de abordar los riesgos vinculados a determinados usos de esta nueva tecnología. En ese mismo año, el Parlamento Europeo estableció una comisión para examinar el impacto de la tecnología (Comisión Especial sobre Inteligencia Artificial en la Era Digital, P9_TA(2020)0162) y pidió una normativa sobre inteligencia artificial antropocéntrica.

²³ Resolución del Parlamento Europeo, de 20 de octubre de 2020, sobre los derechos de propiedad intelectual para el desarrollo de las tecnologías relativas a la inteligencia artificial, 2020/2015(INI).

2024/1689, la mención a los grandes modelos de inteligencia artificial generativa queda limitada a un par de párrafos de la exposición de motivos en el que se afirma que son ejemplos típicos de modelos de inteligencia artificial y se describe el complejo escenario para los artistas y autores en general y a su mención en el anexo. En todo caso, el Reglamento se aplica a todos los modelos de inteligencia artificial que se regulan en su capítulo V²⁴.

En cualquier caso, el Reglamento (UE) 2024/1689 no solo establece las normas armonizadas para la introducción en el mercado, la puesta en servicio y la utilización de sistemas de IA (y por lo tanto de inteligencia artificial generativa) en la Unión sino que recoge las prohibiciones de determinadas prácticas de IA y los requisitos específicos para los sistemas de IA de alto riesgo y obligaciones para los operadores de dichos sistemas; asimismo define las normas armonizadas de transparencia aplicables a determinados sistemas de IA, las normas armonizadas para la introducción en el mercado de modelos de IA de uso general y las normas sobre el seguimiento del mercado, la vigilancia del mercado, la gobernanza y la garantía del cumplimiento. La norma europea contempla también un conjunto de medidas en apoyo de la innovación, prestando especial atención a las pymes, incluidas las empresas emergentes (Guillem Carrau, 2024, p. 84).

Todos los sistemas de IA de alto riesgo deben ser evaluados antes de su comercialización y a lo largo de su ciclo de vida. Ya se tuvo ocasión de desgranar (Guillem Carrau, 2024, p. 105) que la legislación europea distingue sistemas de inteligencia artificial de riesgo inaceptable, de alto riesgo y de riesgo limitado. Los primeros, los identificados como sistemas de inteligencia artificial de riesgo inaceptable, son los que se consideran una amenaza para las personas y deben ser prohibidos²⁵. Los segundos, los de alto riesgo, son

²⁴ El apartado 99 de su Exposición de motivos afirma que *los grandes modelos de IA generativa son un ejemplo típico de un modelo de IA de uso general, ya que permiten la generación flexible de contenidos, por ejemplo, en formato de texto, audio, imágenes o vídeo, que pueden adaptarse fácilmente a una amplia gama de tareas diferenciadas. Así, en particular en el apartado 105, se incide en que los diversos aspectos relacionados con la protección de la propiedad intelectual al afirmar que los modelos de IA de uso general, en particular los grandes modelos de IA generativos, capaces de generar texto, imágenes y otros contenidos, presentan unas oportunidades de innovación únicas, pero también representan un desafío para los artistas, autores y demás creadores y para la manera en que se crea, distribuye, utiliza y consume su contenido creativo. El desarrollo y el entrenamiento de estos modelos requiere acceder a grandes cantidades de texto, imágenes, vídeos y otros datos. Las técnicas de prospección de textos y datos pueden utilizarse ampliamente en este contexto para la recuperación y el análisis de tales contenidos, que pueden estar protegidos por derechos de autor y derechos afines. Todo uso de contenidos protegidos por derechos de autor requiere la autorización del titular de los derechos de que se trate, salvo que se apliquen las excepciones y limitaciones pertinentes en materia de derechos de autor. La Directiva (UE) 2019/790 introdujo excepciones y limitaciones que permiten reproducciones y extracciones de obras y otras prestaciones con fines de prospección de textos y datos en determinadas circunstancias. Con arreglo a estas normas, los titulares de derechos pueden optar por reservarse sus derechos en relación con sus obras u otras prestaciones para evitar la prospección de textos y datos, salvo que su finalidad sea la investigación científica. Cuando el titular del derecho se haya reservado de manera adecuada el derecho de exclusión, los proveedores de modelos de IA de uso general deben obtener su autorización para llevar a cabo una prospección de textos y datos con dichas obras.*

²⁵ Dos ejemplos sería la manipulación cognitiva del comportamiento de personas o grupos vulnerables específicos: por ejemplo, juguetes activados por voz que fomentan comportamientos peligrosos

los sistemas de IA que afecten negativamente a la seguridad o a los derechos fundamentales y se dividirán en dos categorías. El primer grupo abarca los sistemas de IA que se utilicen en productos sujetos a la legislación de la UE sobre seguridad de los productos²⁶. El segundo grupo comprende los sistemas de IA pertenecientes a ocho ámbitos específicos que deberán registrarse en una base de datos de la UE: identificación biométrica y categorización de personas físicas; gestión y explotación de infraestructuras críticas; educación y formación profesional; empleo, gestión de trabajadores y acceso al autoempleo; acceso y disfrute de servicios privados esenciales y servicios y prestaciones públicas; aplicación de la ley; gestión de la migración, el asilo y el control de fronteras y la asistencia en la interpretación jurídica y aplicación de la ley²⁷. Por último, la tercera categoría, los sistemas de IA de riesgo limitado, deben cumplir unos requisitos mínimos de transparencia que permitan a los usuarios tomar decisiones con conocimiento de causa. Tras interactuar con las aplicaciones, el usuario puede decidir si desea seguir utilizándolas. En este sentido, la norma contempla una serie de requisitos de transparencia para la inteligencia artificial generativa, tales como revelar que el contenido ha sido generado por IA, diseñar el modelo para evitar que genere contenidos ilegales y publicar resúmenes de los datos protegidos por derechos de autor utilizados para el entrenamiento.

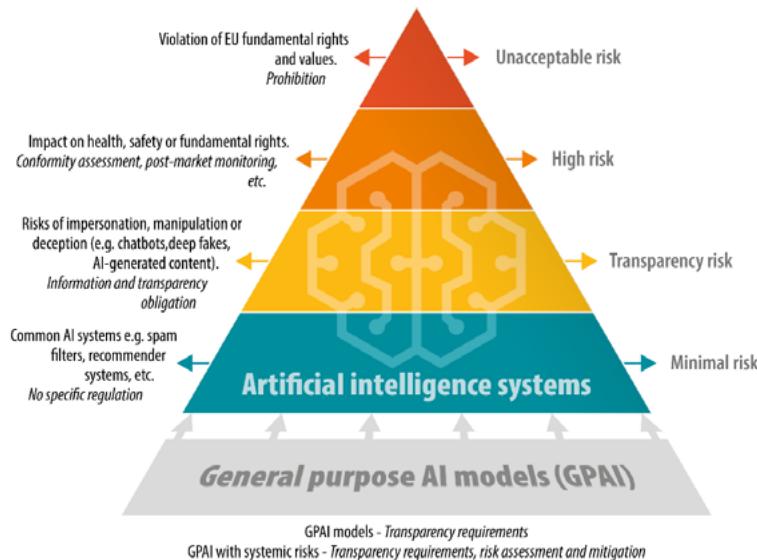
Además, en este contexto, la inteligencia artificial de cualquier categoría, incluidos los modelos extensos de lenguaje, puede caer en lo que se define como de riesgo inaceptable a todos los sistemas de IA considerados una clara amenaza para la seguridad, los medios de subsistencia y los derechos de las personas están prohibidos²⁸.

en los niños; o la puntuación social: clasificación de personas en función de su comportamiento, estatus socioeconómico o características personales.

²⁶ Esto incluye juguetes, aviación, automóviles, dispositivos médicos y ascensores. Véase artículo 16 y ss. del Reglamento (UE) 2024/1689.

²⁷ Artículos 50 y ss. del Reglamento (UE) 2024/1689.

²⁸ Véase <https://digital-strategy.ec.europa.eu/es>. Así, reconoce como de alto riesgo a los casos de uso de la IA que pueden plantear riesgos graves para la salud, la seguridad o los derechos fundamentales y, en particular, los componentes de seguridad de la IA en infraestructuras críticas (por ejemplo, el transporte), cuyo fallo podría poner en peligro la vida y la salud de los ciudadanos; las soluciones de IA utilizadas en instituciones educativas, que pueden determinar el acceso a la educación y el curso de la vida profesional de una persona (por ejemplo, puntuación de los exámenes); los componentes de seguridad de los productos basados en IA (por ejemplo, aplicación de IA en cirugía asistida por robot); las herramientas de IA para el empleo, la gestión de los trabajadores y el acceso al trabajo por cuenta propia (por ejemplo, software de clasificación de CV para la contratación); determinados casos de uso de la IA utilizados para dar acceso a servicios públicos y privados esenciales (por ejemplo, calificación crediticia que niega a los ciudadanos la oportunidad de obtener un préstamo); los sistemas de IA utilizados para la identificación biométrica remota, el reconocimiento de emociones y la categorización biométrica (por ejemplo, un sistema de IA para identificar retroactivamente a un ladrón de tiendas) y los casos de uso de la IA en la aplicación de la ley que pueden interferir con los derechos fundamentales de las personas (por ejemplo, evaluación de la fiabilidad de las pruebas) así como los casos de uso de la IA en la gestión de la migración, el asilo y el control fronterizo (por ejemplo, examen automatizado de las solicitudes de visado), así



Fuente: Tambiama Madiega, «Artificial Intelligence Act: Briefing—EU Legislation in Progress», European Parliamentary Research Service, September 2024, [https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI\(2021\)698792_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf)

Junto al Reglamento, la Comisión ha aprobado unas directrices sobre la definición de sistemas de IA²⁹ y unas directrices sobre las prácticas prohibidas de inteligencia artificial (IA), en las que se menciona la posibilidad de que los sistemas de inteligencia artificial generativa presenten incidentalmente información falsa o engañosa y alucinaciones. En estos casos, la Comisión apunta que el proveedor del sistema debe informar a los usuarios sobre las limitaciones del sistema e integrar medidas adecuadas en el sistema para minimizar estos problemas siempre que el sistema no se utilice en contextos sensibles (por ejemplo, la sanidad, la educación y las elecciones) donde puede haber consecuencias dañinas³⁰.

Por último, en el marco de los países del Espacio Europeo de Investigación, la Comisión ha presentado un conjunto de directrices para apoyar a la comunidad investigadora europea en su uso responsable de la inteligencia

como las soluciones de IA utilizadas en la administración de justicia y los procesos democráticos (por ejemplo, soluciones de IA para preparar resoluciones judiciales).

²⁹ Con estas directrices, la Comisión pretende ayudar a los proveedores y otras personas pertinentes a determinar si un sistema de software constituye un sistema de IA para facilitar la aplicación efectiva de las normas. Véase: Comisión Europea (2025): Communication from the Commission - Commission Guidelines on the definition of an artificial intelligence system established by Regulation (EU) 2024/1689 (AI Act), Brussels, 6.2.2025 C(2025) 924 final accesibles en <https://digital-strategy.ec.europa.eu/es/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application>

³⁰ Comisión Europea (2025): Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act) Brussels, 4.2.2025, (2025) 884 final, accesibles en <https://digital-strategy.ec.europa.eu/es/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act>

artificial generativa³¹. De hecho, la norma europea elabora un listado de usos prohibidos de dichas herramientas en la UE³² y unas reglas para clasificar los sistemas de IA como sistemas de alto riesgo³³.

En sus consideraciones sobre el impacto del Reglamento en los Parlamentos, De Alba Bastarrechea (2025, p. 180) desgrana, sobre la base del principio de responsabilidad proactiva, *un conjunto de obligaciones complejas —evaluaciones de riesgo, supervisión humana, transparencia, documentación técnica, gobernanza de datos— que obligan a los Parlamentos a desarrollar capacidades organizativas y técnicas especializadas*. Asimismo, destaca la necesidad de una actuación *coordinada e interdisciplinaria entre los cuerpos técnicos parlamentarios: los letrados deben interpretar y aplicar el marco normativo europeo e interno, estableciendo límites funcionales a los usos institucionales de la IA; los informáticos parlamentarios deben garantizar la seguridad, trazabilidad y explicación de los sistemas, alineándolos con los principios constitucionales y los archiveros-documentalistas deben asegurar la preservación, clasificación y accesibilidad de los resultados generados por la IA, evitando su opacidad*.

En consecuencia, es evidente que este Reglamento europeo es de aplicación al uso de la IA generativa en el Parlamento o por sus señorías o sus grupos parlamentarios porque se aplica a cualquier operador que ofrezca servicios con tecnología de inteligencia artificial a las Instituciones parlamentarias y a estas mismas si desarrollan programas o servicios sobre la base de sistemas de inteligencia artificial (Guillem Carrau, 2024, p. 93).

Procede pues entender que el principio de legalidad implica que cualquier puesta en práctica de modelos extensos de lenguaje en el ámbito parlamentario tendrá que tener en cuenta las disposiciones del Reglamento (UE) 2024/1689 y, sobre todo, tenerlo de referencia para la definición de los riesgos en las fases de elaboración de modelos extensos de lenguaje para uso parlamentario o de utilización de estos en las actividades parlamentarias y, entre ellas, la legislativa.

IV. LA INCORPORACIÓN RESPONSABLE DE LOS MODELOS EXTENSOS DE LENGUAJE EN EL ÁMBITO PARLAMENTARIO

Los efectos positivos que se describen de la incorporación de los modelos extensos de lenguaje al ámbito parlamentario giran alrededor de la mejora de la eficiencia y la productividad en los procesos de elaboración de las leyes. La IA puede ayudar a los legisladores a analizar grandes cantidades de datos y a identificar patrones y relaciones que pueden no ser evidentes para los humanos, con una información más precisa y actualizada;

³¹ Abril 2025, disponibles en https://research-and-innovation.ec.europa.eu/research-area/industrial-research-and-innovation/artificial-intelligence-ai-science_en

³² Artículo 5.1, Reglamento (UE) 2024/1689.

³³ Artículo 6, Reglamento (UE) 2024/1689.

asimismo, otro efecto positivo consiste en la mejora de la transparencia y la rendición de cuentas porque la puesta en práctica de modelos extensos de lenguaje puede ayudar a los legisladores a mejorar la transparencia y la rendición de cuentas, al proporcionar información más precisa y actualizada sobre las decisiones tomadas.

Entre las fortalezas que pueden aportar al funcionamiento de las Cámaras legislativas se citan la automatización y eficiencia en procedimientos, el aprendizaje automático, la capacidad de procesar gran cantidad de datos y en definitiva la mejora en la toma de decisiones. Asimismo, se identifican un conjunto de amenazas relacionadas con la dependencia excesiva que se genera respecto a éstos, los principios éticos y la necesidad de hacer evolucionar las capacidades de los empleados públicos. Asimismo, estas novedades no están exentas de puntos débiles como son los sesgos y riesgos de seguridad y de privacidad; la aparición de las llamadas cajas negras en torno a los algoritmos o el identificado como déficit de explicabilidad o de claridad, así como las alucinaciones o los costes elevados tanto económicos como medioambientales.

Sobre la base de las prácticas parlamentarias actuales y las investigaciones en curso que lleva a cabo el Centro para la Innovación en el Parlamento de la Unión Interparlamentaria, se ha animado a los Parlamentos a que *empiecen a experimentar con la inteligencia artificial generativa, por ejemplo, para ahorrar tiempo con tareas como elaborar resúmenes de textos y crear registros de los debates parlamentarios. Ese tipo de experimentación debería llevarse a cabo en un «entorno de laboratorio» seguro antes de implantar la tecnología de una forma más generalizada. Sin embargo, los Parlamentos deben extremar las precauciones a la hora de introducir la IA generativa en los sistemas legislativos esenciales en esta fase. Por consiguiente, se requieren unos elevados niveles de confianza en la tecnología y unos sólidos procesos de gobernanza antes de plantearse dar ese tipo de paso*³⁴.

En contexto, se sugieren a continuación diversas soluciones técnicas y estrategias que se demuestran clave para una mejor adaptación de estas herramientas de inteligencia artificial generativa al contexto parlamentario. Existen cuatro elementos técnicos asociados al uso de la inteligencia artificial generativa en el ámbito parlamentario que son de gran interés y que merecen un análisis separado: el encuadre en los procesos de toma de decisión parlamentario por decisiones estructuradas y consistentes de las Mesas de las Cámaras; las normas de utilización de las cajas blancas que ayudan a explicar cómo funciona el programa informático; la definición e identificación de las audiencias de los modelos porque la identificación de la mismas y, en particular, de la interacción entre *laypersons* (sus usuarios) y *domain experts* (los técnicos informáticos) porque es importante para garantizar que los modelos sean desarrollados y utilizados de manera efectiva y responsable; y por último, la posibilidad de determinar con carácter previo cuánto influye cada

³⁴ Inter-parliamentary Union (2024): *Guides for AI in parliaments* en <https://www.ipu.org/ai-guidelines>

elemento del programa informático en la predicción o clasificación realizada por el modelo y en las opciones que presenta como soluciones o aquellas que descarta y no presenta, porque sirve para identificar fácilmente las características del modelo.

4.1. La incorporación a las tramitaciones de textos legislativos y el rol de las Mesas como órgano de gobierno interior

La incorporación de los modelos extensos de lenguaje al procedimiento legislativo supone una oportunidad para el desarrollo de iniciativas legislativas de mayor calidad siempre que dicha incorporación sea pilotada de una manera estructurada y consistente.

El uso de los modelos de lenguaje masivos permite que dichas herramientas se entrenen y crezcan sin supervisión, auto aprendiendo, para resumir documentos y completar frases o, en el caso de la elaboración de las normas, incluso preceptos de las propuestas legislativas o de las enmiendas a dichas propuestas legislativas (Guillem Carrau, 2024, p. 99).

Con carácter preliminar, cabe plantearse si es necesario o no proceder a la modificación de los reglamentos parlamentarios para la inclusión expresa de la posibilidad de utilizar este tipo de herramientas en los procedimientos que corresponda. En principio, no parece necesario mientras se cumplan en el desarrollo y planificación de la puesta en práctica de estos los requisitos y prescripciones establecidas por el Reglamento (UE) 2024/1689 y siempre que se realice el esfuerzo correspondiente en términos de transparencia y explicabilidad de los mismos de cara a los actores parlamentarios y a la ciudadanía.

Con ánimo descriptivo, en el marco de los reglamentos parlamentarios, se pueden enumerar diversos campos de trabajo parlamentario en los que poner en desarrollo el uso de este tipo de modelos de lenguaje con ánimo de mejorar la calidad de la actividad parlamentaria.

En primer lugar, los reglamentos parlamentarios podrán servir de base para permitir el uso de modelos extensos de lenguaje para analizar y procesar información y documentación relacionada con las iniciativas legislativas, siempre que se garanticen los principios de seguridad, veracidad y accesibilidad de la información con carácter transparente a los distintos actores parlamentarios implicados en el proceso de producción legislativa.

En segundo lugar, los Reglamentos podrían permitir el uso de modelos extensos de lenguaje para analizar y evaluar información relacionada con los procedimientos legislativos en general, siempre que se solicite la opinión de expertos y organismos especializados. Tanto en la fase de participación ciudadana, como en la de trabajo en ponencia, el uso de este tipo de herramientas puede ser de gran impacto para la mejora de los textos en tramitación parlamentaria. Asimismo, se podría permitir el uso de modelos extensos

de lenguaje para analizar y evaluar las enmiendas presentadas a los textos en tramitación parlamentaria.

En tercer lugar, se podría permitir el uso de modelos extensos de lenguaje para analizar y evaluar la información intercambiada con otras instituciones y organismos, siempre que se garantice la seguridad y la accesibilidad de la información. Consideremos una serie de tareas cotidianas de carácter administrativo-parlamentario que se llevan a cabo en el Parlamento: realizar investigaciones, escribir documentos y discursos, gestionar enmiendas y redactar legislación³⁵.

Por ejemplo, en lo que concierne a los procesos legislativos, la Cámara de los Representantes de Estados Unidos ha puesto en funcionamiento una herramienta de AI para analizar automáticamente las diferencias entre propuestas legislativas, enmiendas y textos legales en vigor. La herramienta es capaz de interpretar citas legales cruzadas y las modificaciones que en otros textos legales tendrían que operarse en consecuencia con el nuevo precepto legal³⁶.

Respecto a la incorporación estructurada y consistente de los modelos extensos de lenguaje o cualquier otro tipo de IA generativa al entorno parlamentario, es necesario precisar que no existe mejor alternativa al hecho de que el proceso de desarrollo o de uso de las aplicaciones basadas en inteligencia artificial generativa sea impulsado y liderado desde la Mesa de la institución parlamentaria.

En primer lugar, la Mesa ha de definir los límites que se quieran establecer a priori con carácter adicional a los establecidos por el legislador europeo respecto a los riesgos inaceptables y los sujetos a limitación. Tendrá que ser la Mesa de la Institución parlamentaria, asesorada por los órganos técnicos que procedan y la Secretaría General, la que defina las líneas rojas que delimiten el uso de la tecnología sobre la base de un análisis de riesgos para evitar su uso en áreas donde el riesgo sea alto o su uso con especial atención en áreas donde los controles sean indispensables y aquellas donde, al ser los riesgos menores, puedan ser empleadas más libremente.

Asimismo, como órgano de gobierno, la Mesa deberá hacer público los sistemas o herramientas de inteligencia artificial que se emplean en la actividad parlamentaria y en las relaciones con la ciudadanía, el riesgo asociado a cada uno de los casos e identificar los contenidos o informaciones generadas por dichos sistemas de manera que puedan hacerse públicos y accesibles (Guillem Carrau, 2024, p. 106).

En este contexto, son de utilidad las consideraciones que se desarrollan en los puntos subsiguientes en torno a la preferencia sobre unas soluciones técnicas respecto a otras: los modelos de caja blanca frente a otros, la defini-

³⁵ Inter-parliamentary Union (2024): Guides for AI in parliaments en <https://www.ipu.org/ai-guidelines>, p. 3.

³⁶ US Congress (2025): Regulating Artificial Intelligence: U.S. and International Approaches and Considerations for Congress June 4, 2025 Congressional Research Service, accessible en <https://crsreports.congress.gov/R48555>.

ción extensa de la audiencia y su implicación en la definición de los modelos, y la identificación clara de las características y escenarios alternativos en cada uno de los modelos que se vayan a desarrollar.

Por último, la Mesa de la Cámara, como órgano de gobierno interior, deberá planear y desarrollar de manera consistente no solo el aterrizaje de estos procesos de manera estructurada sino planificar y dotar de recursos necesarios los planes de capacitación y formación de los distintos actores parlamentarios que vayan a interactuar con los modelos extensos de lenguaje, incluyendo el personal al servicio de la institución parlamentaria.

4.2. Las cajas blancas y la explicabilidad de los modelos

Los modelos extensos de lenguaje producen soluciones ante los *prompts* que se les formulan. Estas soluciones o decisiones propuestas se definen como decisiones fiduciarias; esto es, se hace necesario comprobar la decisión para confirmar su credibilidad y explicar la importancia de las características tanto por usuarios finales, por ejemplo, la ciudadanía, y expertos, por ejemplo, letrados o técnicos en la materia.

Además, la motivación de la decisión será también necesaria para conocer y cuestionar la decisión y explicar la importancia de las características y los escenarios alternativos. Esta motivación pasará el filtro del usuario final.

En este sentido, como refuerzo de seguridad y responsabilidad, los modelos extensos de lenguaje que se empleen en relación con la función legislativa deberán mostrar la ausencia de errores con explicaciones globales o modelos de cajas blancas³⁷ y esto deberá ser certificado por los expertos en machine Learning o aprendizaje automático.

Se trata de desarrollar modelos extensos de lenguaje que decidan cómo asignar recursos para maximizar el bienestar de una persona o comunidad y que tomen decisiones sobre cómo proteger la privacidad y la seguridad de los datos de un usuario. Asimismo, deberán ser capaces de equilibrar los objetivos en conflicto de diferentes partes interesadas, como la rentabilidad y la sostenibilidad ambiental.

En el contexto de los modelos extensos de lenguaje, los modelos de caja blanca están llamados a adquirir mucho protagonismo (Yeung *et al.*, 2022). La razón de ello es que los modelos de aprendizaje automático de «caja blanca» son diseñados para ser transparentes y comprensibles, es decir, que permiten a los usuarios entender cómo funcionan y cómo toman decisiones. Básicamente, son transparentes en relación con los datos usados para entrenar el algoritmo, su gravedad y su relevancia para la solución y la lógica empleada en proponerla. Un ejemplo de ello es el ojo de halcón

³⁷ Sobre estos modelos en general, véase Loyola-González, O., Black-box vs. white-box: understanding their advantages and weaknesses from a practical point of view. IEEE Access. 2019;7:154096–113.

utilizado para las competiciones deportivas, sustituyendo las decisiones de los jueces deportivos.

En síntesis, los beneficios de la aplicación de inteligencia artificial de caja blanca es que está diseñada para una comprensión reforzada, permitiendo una supervisión humana efectiva e intervención cuando sea necesario evitar o corregir el modelo lógico. Los modelos de caja blanca suelen tener las siguientes características: interpretabilidad, transparencia, explicabilidad y flexibilidad. De este modo, los modelos de caja blanca: están diseñados para ser fáciles de entender y interpretar, lo que permite a los usuarios comprender cómo funcionan y cómo toman decisiones; proporcionan información clara y concisa sobre cómo se toman las decisiones, lo que permite a los usuarios evaluar la lógica y la coherencia del modelo; pueden proporcionar explicaciones claras y concisas sobre por qué se tomaron ciertas decisiones, lo que puede ser útil para identificar posibles errores o sesgos; y, por último, pueden ser modificados y ajustados para adaptarse a diferentes escenarios y requisitos (Kouroutakis, 2024).

La doctrina apunta que algunos ejemplos de modelos de caja blanca incluyen: modelos de árboles de decisión; modelos de regresión lineal; modelos de lógica difusa y modelos de redes neuronales interpretativas que son un tipo de modelo de caja blanca que utiliza técnicas de visualización y explicación para hacer que las redes neuronales sean más comprensibles. Entre las ventajas de los modelos de caja blanca, se incluyen la confianza porque pueden generar mayor confianza en los usuarios, ya que permiten entender cómo funcionan y cómo toman decisiones; la interpretatividad porque pueden mejorar la interpretación de los resultados, lo que puede ser útil para identificar posibles errores o sesgos; y la flexibilidad porque pueden ser modificados y ajustados para adaptarse a diferentes escenarios y requisitos (Wachter *et al.*, 2022).

Entre sus inconvenientes se relacionan la complejidad que les caracteriza respecto a los de caja negra, lo que implica la necesidad de más recursos de mantenimiento; las limitaciones en términos de capacidad para manejar datos complejos y no lineales. Asimismo, tienen límites para solucionar problemas más complejos o su corrección predictiva podría ser menor ante datos no lineales, interconectados o heterogéneos (Kouroutakis, 2024).

En todo caso, es necesario precisar que la explicabilidad se exige a todos los modelos, blancos o negros, extensos de lenguaje. La calidad de las explicaciones es clave y debe ser clara, apropiada y suficiente que no crea dudas sobre la decisión tomada con el riesgo de la dificultad de que las explicaciones funcionen correctamente porque habrá modelos que aporten soluciones ambiguas o inconsistentes (Bordt *et al.*, 2022).

Por ello, parece que son los modelos más adecuados para la implementación en el desarrollo de tareas legislativas. La caja blanca va a permitir que se pueda explicar a la ciudadanía y a los otros miembros del Parlamento las soluciones propuestas por el mismo. Estas soluciones pueden ser desde textos legislativos alternativos al que se encuentra en trámite, enmiendas al mismo

desde una determinada perspectiva ideológica o técnica, la realización de compilaciones legislativas, la búsqueda de soluciones de derecho comparado, etc. En este sentido, se entiende claramente que la caja blanca permitirá transparencia y trazabilidad sobre por qué se tomaron ciertas decisiones; la identificación de errores, alucinaciones y sesgos; y, por último, como hemos visto, pueden ser modificados y ajustados para adaptarse a diferentes escenarios y requisitos de la arena parlamentaria.

4.3. La definición de la audiencia o público de los modelos extensos de lenguaje

Otro elemento importante en el marco de la utilización de los modelos extensos de lenguaje en la función legislativa es la audiencia o público de estos programas informáticos. Se tiende a afirmar que, en el contexto de la Inteligencia Artificial General, es necesario identificar dos colectivos singulares: los usuarios (conocidos en inglés como «*laypersons*») y los expertos (conocidos en inglés como «*domain experts*»); que se refieren a dos grupos de personas con diferentes niveles de conocimiento y experiencia en relación con la tecnología de inteligencia artificial generativa y su aplicación en un dominio específico.

Por un lado, los «no expertos» o «usuarios finales» son personas que no tienen un conocimiento profundo o especializado en la tecnología de inteligencia artificial generativa ni en el dominio específico en el que se aplica. Pueden ser usuarios finales que interactúan con los modelos extensos de lenguaje, pero no necesariamente entienden cómo funcionan internamente. Pueden tener una comprensión básica de la tecnología, pero no tienen la capacidad de evaluar críticamente sus limitaciones o debilidades. Ejemplos de modelos extensos de lenguaje podrían incluir: los usuarios de estas aplicaciones en sus teléfonos inteligentes o asistentes virtuales; los clientes de estos servicios que no tienen un conocimiento técnico profundo; y las personas que utilizan herramientas de este tipo en su vida diaria, como sistemas de recomendación o asistentes de escritorio.

Por otro lado, los «expertos en el dominio» o «especialistas» son personas que tienen un conocimiento profundo y especializado en un dominio específico, como la medicina, la finanza, la ingeniería, etc. Pueden tener un conocimiento básico o intermedio de la tecnología de GenAI, pero su experiencia y peritaje se centran en el dominio específico en el que se aplica. Pueden evaluar críticamente las fortalezas y debilidades de los modelos extensos de lenguaje en su dominio específico y pueden identificar oportunidades de mejora. Ejemplos de expertos en el dominio podrían incluir a los Médicos que utilizan sistemas de inteligencia artificial generativa para diagnosticar enfermedades o desarrollar tratamientos personalizados; los ingenieros que utilizan herramientas de este tipo para diseñar y optimizar sistemas comple-

jos; los economistas que utilizan modelos basados en inteligencia artificial generativa para analizar y predecir tendencias económicas.

En el contexto de los modelos extensos de lenguaje y su uso en la función legislativa, la interacción entre usuarios y expertos es importante para garantizar que los modelos sean desarrollados y utilizados de manera efectiva y responsable en los ámbitos parlamentarios. Los expertos pueden proporcionar retroalimentación y orientación a los desarrolladores del modelo, mientras que los usuarios pueden proporcionar una perspectiva más amplia y general sobre cómo los modelos pueden ser utilizados y beneficiar a la sociedad en general.

Por ello, como se ha anticipado, las Mesas de las instituciones parlamentarias deberán definir las audiencias de cada modelo extenso de lenguaje que se vaya a incorporar a los trabajos parlamentarios con carácter previo para que se puedan incorporar a su proceso de creación, desarrollo y puesta en práctica.

4.4. Las características y los escenarios alternativos

La identificación de las características y de los escenarios alternativos se convierten en un elemento fundamental para la utilización responsable y ética de los modelos extensos de lenguaje en la actividad parlamentaria y, en particular, en la función legislativa. Así, la posibilidad de determinar con carácter previo cuánto influye cada característica en la predicción o clasificación realizada por el modelo sirve para identificar hasta dónde puede ser de utilidad el modelo en cuestión.

La importancia de las características puede ayudar a identificar las variables que tienen un impacto significativo en la predicción o clasificación, lo que puede ser útil para reducir la dimensionalidad del conjunto de datos y mejorar la eficiencia del modelo. De este modo, también puede ayudar a identificar las variables que no tienen un impacto significativo en la predicción o clasificación, lo que puede ser útil para eliminarlas y reducir el ruido en el conjunto de datos. Todo ello redundaría en una mejora de la interpretación del modelo al proporcionar información de valor añadido sobre cómo funciona el modelo y cómo se relacionan las variables entre sí, lo que puede ser útil para mejorar la interpretación y la confianza en el modelo.

De igual modo, en el contexto de la Inteligencia Artificial General, los escenarios alternativos, o también llamados contrafácticos, se refieren a la idea de evaluar cómo habría sido el resultado de un modelo de aprendizaje automático si se hubieran utilizado diferentes datos de entrada, parámetros o condiciones. En otras palabras, los contrafácticos permiten explorar qué habría pasado si se hubieran tomado decisiones diferentes o si se hubieran utilizado diferentes datos y facilitan la evaluación de la robustez del modelo en diferentes escenarios, lo que puede ser útil para identificar posibles de-

bilidades o vulnerabilidades. Asimismo, los contrafácticos pueden ayudar a analizar la causalidad entre las variables y los resultados e identificar los sesgos y posibles discriminaciones, lo que puede ser útil para entender mejor cómo funciona el modelo y mejorar la interpretación y confianza en el modelo extenso de lenguaje.

Ambos tipos de elementos, características y contrafácticos, son fundamentales para conocer técnicamente, las posibilidades que ofrecen la puesta en práctica de estos modelos extensos de lenguaje en el contexto de las tareas propias de la actividad legislativa; como hemos descrito anteriormente, en la redacción de artículos de normas, enmiendas a los mismos, búsquedas de alternativas en derecho comparado o ejercicios de compilación legislativa. Por ello, las Mesas de las Instituciones parlamentarias tienen la posibilidad de definir estos elementos para cada uno de las herramientas de IA generativa que acuerden incorporar en los entornos parlamentarios. Esta definición es clave para incrementar la confianza de los actores parlamentarios y la ciudadanía en la capacidad de emplear estos nuevos instrumentos de trabajo para mejorar la calidad de las leyes.

V. CONCLUSIONES

Las tecnologías como la inteligencia artificial generativa, con su capacidad de crear contenido basado en grandes cantidades de datos, promete eficacia y productividad y, potencialmente, muchos cambios y transformaciones en los Parlamentos, que no tiene parangón con las mejoras técnicas aportadas por las nuevas tecnologías en las últimas décadas. Muchos Parlamentos utilizan habitualmente la inteligencia artificial (IA). Algunos ejemplos son los antivirus, las aplicaciones de protección contra el correo electrónico no solicitado y los programas de ciberseguridad. Estos sistemas utilizan la IA para adoptar miles de decisiones —a menudo de forma constante— cada día.

Toda estrategia de desarrollo de la inteligencia artificial, pese a los esfuerzos de planificación y de preparación, presenta desafíos y riesgos que no sólo con planificación estratégica se pueden resolver. Como la mayoría de los asuntos tecnológicos, el desarrollo de la inteligencia artificial exige opciones políticas reales y estrategias de implementación firmes, diseñadas para maximizar beneficios y minimizar los resultados negativos o los riesgos asociados a su puesta en práctica relacionados con la seguridad, la privacidad y la transparencia de los sistemas de inteligencia artificial (Guillem Carrau, 2024, p. 101).

Se supone que cualquier programa institucional deberá diseñarse teniendo en cuenta diversas etapas definidas en función de riesgos y oportunidades y el respeto a los principios de igualdad, veracidad, privacidad, responsabilidad y legalidad. En todo lo relacionado con la inteligencia artificial generati-

va y, en particular con los modelos extensos de lenguaje, no se puede perder de vista la importancia de supervisión humana y la transparencia y explicabilidad de estas nuevas herramientas de trabajo.

En primer lugar, se recomienda desarrollar un marco global que refleje los principios éticos del Parlamento en relación con los modelos extensos de lenguaje y la IA en general e identifique los niveles de riesgo del uso de estos conforme al Reglamento (UE) 2024/1689. Por ejemplo, como se ha descrito anteriormente, la apuesta debe ser el uso de modelos extensos de lenguaje que no favorezcan la obtención de resultados o soluciones sexistas o discriminatorias en razón de edad, sexo, raza, origen u otros motivos de discriminación frente a aquellos que no lo hagan.

En segundo lugar, se propone establecer protocolos para asegurar transparencia y trazabilidad en todo lo relacionado con la AI y analizar y mitigar los sesgos de los sistemas de inteligencia artificial. Por ejemplo, su desarrollo debe venir acompañado de medidas de discriminación positiva para, por un lado, equilibrar los posibles sesgos que contengan, y por otro, facilitar acceso a los mismos para personas con distintas capacidades.

En tercer lugar, se apunta la necesidad de desarrollar una estrategia de comunicación para informar a la ciudadana sobre el uso parlamentario de la IA generativa y del modo que se van a tratar las informaciones y datos que comparten a través de estos sistemas con la Institución parlamentaria. Por ejemplo, esta necesidad imperativa se verbaliza en torno a la explicabilidad del modelo y de su régimen de privacidad de datos y, a su vez, por la preferencia por la adopción de modelos de caja blanca y la posibilidad de explicar sus características y sus escenarios alternativos, en el sentido expuesto en estas líneas.

En cuarto lugar, es clave trazar estrategias de mitigación que permitan una experimentación segura con la IA, paso a paso, evaluando riesgos y adoptando directrices para poder introducir la IA en el Parlamento en el marco de su cultura organizativa. Es necesario que la ciudadanía y los actores parlamentarios sean conscientes de que existe una supervisión humana en su diseño y en el seguimiento de los resultados aportados por su utilización en las distintas funciones parlamentarias.

En quinto lugar, será fundamental la colaboración con otros Parlamentos y expertos externos para compartir conocimiento y análisis de procedimientos parlamentarios para su posible integración con los nuevos modelos extensos de lenguaje.

En sexto lugar, la estrategia debe partir con el desarrollo de un marco de ciberseguridad que permita tener garantías para poner en práctica los sistemas de inteligencia artificial y el desarrollo de programas de conocimiento de inteligencia artificial y de literatura de datos para los funcionarios y los miembros de los Parlamentos. Por ejemplo, se propone que las administraciones parlamentarias, por medio de sus servicios informáticos y de los responsables de protección de datos garanticen la privacidad en el uso de las mismas y la compatibilidad del uso de este tipo de herramientas con las medidas de

seguridad implementadas en la arquitectura informática y las aplicaciones empleadas en sede parlamentaria.

En séptimo lugar, será imprescindible crear procesos de gestión de calidad para programas de entrenamiento de los instrumentos que se vayan a utilizar de la inteligencia artificial y establecer un plan de evolución de la gestión para atender cambios de funciones y potenciales impactos en las tareas de funcionarios públicos y miembros de los Parlamentos y capacitarlos para interactuar con los modelos extensos de lenguaje.

Señala la Unión Interparlamentaria que *la inteligencia artificial generativa solo es el comienzo. Llegarán pronto otras formas emergentes de IA que tendrán que ser evaluadas y entendidas en los Parlamentos como la computación neuromórfica, que tiene por objeto copiar el proceso del pensamiento humano digitalmente; y la IA neurosimbólica, que combina las capacidades de aprendizaje estadístico y basado en datos de las redes neuronales con el razonamiento simbólico. Durante los próximos años, los Parlamentos tendrán que evaluar y entender los beneficios y los riesgos que entrañan todas las formas emergentes de IA. La variante generativa es solo el comienzo*³⁸.

Por ello, es de gran importancia sentar las bases organizativas y de procesos para estar preparado para los cambios que se avecinan en el contexto de las sociedades democráticas avanzadas y que el Parlamento siga siendo el centro del debate y decisión política.

VI. BIBLIOGRAFÍA

- Bordt, S.; Finck, M.; Raidl, E., and Luxburg von, U. (2022) *Post-hoc explanations fail to achieve their purpose in adversarial contexts* in (ACM 2022) FAccT '22 891.
- Coeckelbergh, M. (2021). *Ética de la inteligencia artificial*. Madrid: Cátedra.
- Coglianese, Cary and Crum, Colton R., Taking Training Seriously: Human Guidance and Management-Based Regulation of Artificial Intelligence (February 15, 2024). U of Penn Law School, Public Law Research Paper N.º 24-08, Available at SSRN: <https://ssrn.com/abstract=4729072>
- Dahl, M.; Magesh,V.; Suzgun, M. y Ho, D. (2024) Hallucinating Law: Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models, en <https://hai.stanford.edu/news/hallucinating-law-legal-mistakes-large-language-models-are-pervasive>
- De Alba Bastarrechea, E. (2025). La inteligencia artificial en la transformación digital de los Parlamentos: en busca de un modelo ad hoc de gobernanza. *Asamblea. Revista Parlamentaria de la Asamblea de Madrid*, n.º 48, 161–193.
- Fitsilis, F.; Koryzis, D.; Schefbeck, G. Legal Informatics Tools for Evidence-Based Policy Creation in Parliaments. *Int. J. Parliam. Stud.* 2022, 2, 1–25.

³⁸ Véase <https://www.ipu.org/resources/publications/issue-briefs/2024-04/using-generative-ai-in-parliaments>

- Guillem Carrau, J. (2024). Inteligencia artificial y Parlamentos modernos en *Corts:Anuario de derecho parlamentario*, ISSN 1136-3339, n.º 38, 77-114.
- Inter-parliamentary Union (2024): Guides for AI in parliaments en <https://www.ipu.org/ai-guidelines>
- Kouroutakis, A. (2024). Rule of law in the AI era: addressing accountability, and the digital divide. *Discov Artif Intell* 4, 115.
- Loyola-González O. Black-box vs. white-box: understanding their advantages and weaknesses from a practical point of view. *IEEE Access*. 2019;7:154096–113.
- Miller, T. Explanation in artificial intelligence: Insights from the social sciences. *Artif Intell*. 2019;267:1–38.
- Montoya, D. y Rummery, A.: The use of artificial intelligence by government: parliamentary and legal issues, E-brief n.º 02/2020, Parliament of South Wales, en <https://www.parliament.nsw.gov.au/researchpapers/Pages/The-use-of-AI-by-government-parliamentary-and-legal-issues.aspx>
- US Congress (2025): Regulating Artificial Intelligence: U.S. and International Approaches and Considerations for Congress June 4, 2025 Congressional Research Service, en <https://crsreports.congress.gov/R48555>.