REVISTA DE

PRIVACIDAD Y DERECHO DIGITAL

DIRECTOR • D. PABLO GARCÍA MEXÍA

AÑO IX • MAYO-AGOSTO 2024 • NÚMERO 34

ISSN: 2444-5762

RIESGOS DE LOS SISTEMAS DE INTELIGENCIA ARTIFICIAL GENERATIVA (*)

RISKS OF GENERATIVE ARTIFICIAL INTELLIGENCE SYSTEMS AND THE EUROPEAN ARTIFICIAL INTELLIGENCE REGULATION

Por Mª JESÚS JIMÉNEZ LINARES
Profesora Titular de Derecho civil de la Universidad de Granada

(*) Este trabajo se recibió el 11 de junio de 2024 y fue aceptado el 30 de julio.

REVISTA DE
PRIVACIDAD Y DERECHO DIGITAL

RESUMEN

La inteligencia artificial generativa permite la creación automática de contenido "original" en diferentes formatos (texto, audio, vídeo e imágenes). Un contenido que cada vez se acerca más al creado por los humanos. La inteligencia artificial generativa ofrece a la sociedad tantos beneficios como posibles riesgos. Se analizarán algunos como la confusión, la desinformación, los riesgos de los derechos de autor, a la seguridad, a la privacidad, la ciberdelincuencia, al derecho al honor, la intimidad y la propia imagen que pueden lesionarse con los deep fakes, la manipulación, los riesgos en el mercado laboral, los psicológicos, los sesgos...El nuevo Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) acoge la inteligencia artificial generativa dentro de la inteligencia de uso general, teniendo en cuenta la presencia, de posibles riesgos sistémicos y la exigibilidad de la ciberseguridad. Se analizará también la conexión de la inteligencia artificial generativa con la prohibida y la de alto riesgo. El Reglamento muestra especial preocupación ante las "ultrasuplantaciones", deep fakes (donde es difícil distinguir entre lo real y lo irreal), y el necesario conocimiento de su origen artificial cuando las personas físicas interactúen con ellas, con la correspondiente obligación de transparencia. Todo ello, sin olvidar, la necesidad de hacer lo ético jurídico y de la alfabetización.

PALABRAS CLAVE: *Inteligencia artificial generativa, riesgos, desinformación, seguridad, privacidad, ciberseguridad, transparencia, deep fakes, ultrasuplantaciones, riesgos sistémicos, ley de inteligencia artificial.*

ABSTRACT

Generative artificial intelligence allows the automatic creation of "original" content in different formats (text, audio, video and images). Content that is getting closer and closer to that created by humans. Generative artificial intelligence offers society as many benefits as possible risks. Some will be analysed such as confusion, disinformation, copyright risks, security, privacy, cybercrime, the right to honour, privacy and self-image that can be injured by deep fakes, manipulation, risks in the labour market, psychological risks, bias... The new Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) welcomes generative artificial intelligence within general purpose intelligence, considering the presence of possible systemic risks and the enforceability of cybersecurity. The connection of generative artificial intelligence with prohibited and high-risk artificial intelligence will also be analysed. The Regulation shows special concern for "ultra-phishing", deep fakes (where it is difficult to distinguish between the real and the unreal), and the necessary knowledge of their artificial origin when natural persons interact with them, with the corresponding obligation of transparency. All this, without forgetting the need for legal ethics and literacy.

KEY WORDS: *Generative artificial intelligence, risks, disinformation, security, privacy, cybersecurity, transparency, deep fakes, ultra spoofing, systemic risks, artificial intelligence law.*

SUMARIO

I.- INTRODUCCIÓN[1]

En 1984, en la película "los gremlins" aparecieron unas "criaturas" muy especiales. Gizmo, un mogwai, ser adorable, cuyo dueño, Mr.Wing, no pensaba, en principio, venderlo a Rand Peltzer (que lo quería para su hijo Billy) porque suponía una gran responsabilidad. Gizmo acabó generando unos seres totalmente destructivos, como Stripe, asesinos sin escrúpulos. Todo ello provocado por un comportamiento descuidado humano, por el incumplimiento de unas normas mínimas con Gizmo: a la criatura no le gustaba la luz brillante, la luz del sol lo mataría y nunca debían darle agua (ni bañarlo) ni darle de comer después de la medianoche[2].Finalmente, ante el caos que se generó en la sociedad, Mr. Wing volvió a recuperar a Gizmo.

Igual que los gremlins una simple IA que nos divierte generando vídeos, fotos, imágenes o clonando voces, puede llegar a convertirse en la causante de amenazas múltiples y poner en jaque la seguridad nacional, si no se ejerce el uso y el control humano adecuado de la misma y se establece su correcta regulación. La gran paradoja de la IA y en este caso concreto de la IA generativa es ser portadora a la vez de grandes beneficios y riesgos como se verá posteriormente.

1 Este trabajo forma parte del Proyecto de investigación "Nuevos avances en la legislación de transparencia en España: mejoras en la definición del marco regulatorio" (PID 2021-124724NB-100), del que es IP la profesora Ana de Marcos Fernández.

2 WIKIPEDIA, https://es.wikipedia.org/wiki/Gremlins (recuperado el 1 de mayo del 2024).

II.- LA IA GENERATIVA: CONCEPTO, BENEFICIOS Y RIESGOS

El 12 de julio del 2024 se publicó el 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 (la llamada Ley de Inteligencia Artificial, a partir de ahora LIA)[3] centrado y enfocado en la gestión de riesgos. Se clasifica la IA en función de su riesgo potencial e impacto, imponiéndose más obligaciones y control cuanto mayor es el riesgo. Busca una IA fiable, ética, digna de confianza, cuyo eje sea el ser humano. Establece un marco jurídico de normas armonizadas que protegen fuertemente los intereses públicos, como la salud y la seguridad y los derechos fundamentales, particularmente, la democracia, el Estado de Derecho y el medio ambiente. En conexión con la LIA, es necesario fomentar un ecosistema de confianza preocupado por la rendición de cuentas (la responsabilidad), los datos, desarrollo de confianza y despegue, notificación de incidentes, pruebas y garantía y seguridad[4]. Se analizarán los riesgos de la IA generativa y su presencia en la LIA.

3 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) nº 300/2008, (UE) nº 167/2013, (UE) nº 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, Serie L, 12 de julio del 2024. Una visión general sobre el Reglamento se puede ver en: FERNÁNDEZ HERNÁNDEZ, C., et al. "Diez puntos críticos del Reglamento europeo de Inteligencia Artificial", Diario LA LEY, Sección Ciberderecho, nº 85, 28 de junio de 2024 y BARRIO ANDRÉS, M., "Algunos claroscuros en el Reglamento Europeo de Inteligencia Artificial", Diario La Ley, nº 86. Sección Ciberderecho, 30 de julio de 2024 y sobre la conjugación de la innovación y la regulación, GARCÍA MEXÍA, P. "Europa ante el reto de la inteligencia artificial",The objective, 3 de agosto del 2024, (https://theobjective.com/tecnologia/2024-08-03/europa-ante-el-reto-de-la-inteligencia-artificial/) (recuperado el 5 de agosto del 2024).

4 Informe Model AI Governance Framework for Generative AI Fostering a Trusted Ecosystem, de 30 de mayo del 2024 (pág.5). En el Informe se muestra: -La rendición de cuentas: establecer la estructura de incentivos adecuados para que los distintos participantes en el ciclo de vida de desarrollo del sistema de IA sean responsables ante los usuarios finales; -Datos: garantizar la calidad de los datos y abordar los datos de formación potencialmente conflictivos de forma pragmática, y que son fundamentales para el desarrollo de modelos; -Desarrollo de confianza y despliegue: aumentar la transparencia en torno a las medidas

II.1.- EL CONCEPTO DE IA GENERATIVA

El concepto de IA generativa no aparece en la LIA, ni en sus definiciones (art.3 LIA). Si bien, sí define la IA que la engloba, la IA de uso general, así, para la LIA (en su Considerando 99) "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 diferencias[5]". En el documento de debate Generative AI:Implications for Trust and Governance, 4, publicado en junio del 2023, se expresa que la Inteligencia Artificial General (IAG)

básicas de seguridad e higiene basadas en las mejores prácticas del sector, en desarrollo, evaluación y divulgación; -Notificación de incidentes: implantación de un sistema de gestión de incidentes para la notificación y corrección oportunas y mejoras continuas, ya que ningún sistema de IA es infalible; -Pruebas y garantía: proporcionar validación externa y confianza añadida mediante pruebas de terceros, y desarrollar normas comunes de pruebas de IA en aras de la coherencia; -Seguridad: hacer frente a los nuevos vectores de amenaza que surgen gracias a los modelos generativos de IA; -Procedencia de los contenidos: transparencia sobre la procedencia de los contenidos como señales útiles para los usuarios finales;- I+D en seguridad y alienación: acelerar la I+D mediante la cooperación mundial entre los Institutos de seguridad de la IA para mejorar la alienación de los modelos con la intención y los valores humanos; -IA para el bien público: la IA responsable incluye el aprovechamiento de la IA en beneficio del público mediante la democratización del acceso, la mejora de la adopción por parte del sector público, la mejora de las cualificaciones de los trabajadores y el desarrollo sostenible de los sistemas de IA.

5 Esta idea ya aparecía en el art.3.5 del Real Decreto 817/2023, de 8 de noviembre, por el que se establece un entorno controlado de pruebas para el ensayo del cumplimiento de la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial. Posteriormente la LIA, distingue y nos define el "modelo de IA de uso general" (art.3 definición 63) y el "sistema de IA de uso general" (art.3 definición 66) Siendo el primero: "un modelo de IA, también uno entrenado con un gran volumen de datos utilizando autosupervisión a gran escala, que presenta un grado considerable de generalidad y es capaz de realizar de manera competente una gran variedad de tareas distintas, independientemente de la manera en que el modelo se introduzca en el mercado, y que puede integrarse en diversos sistemas o aplicaciones posteriores, excepto los modelos de IA que se utilizan para actividades de investigación, desarrollo o creación de prototipos antes de su introducción en el mercado" y el segundo como" un sistema de IA basado en un modelo de IA de uso general y que puede servir para diversos fines, tanto para su uso directo como para su integración en otros sistemas de IA". Su capítulo V se dedica a "los modelos de IA de uso general", apareciendo en su sección primera las reglas de clasificación (arts.51 y 52 LIA). En el Considerando 97 se incide en la importancia de la distinción entre modelo y sistema de IA de uso general para garantizar la seguridad jurídica.

"se refiere comúnmente a la IA que posee la capacidad de comprender, aprender y realizar una amplia gama de tareas a un nivel que iguala o supera las capacidades humanas. Contrasta con la IA restringida, que solo puede realizar una tarea específica"[6]. La IA generativa, añade el documento, son modelos generativos que aprenden la distribución subyacente de los datos y pueden generar nuevos contenidos (literatura, audio, vídeos) a partir de esta distribución aprendida. Concepto que queda matizado en el Informe Model AI Governance Framework for Generative AI Fostering a Trusted Ecosystem, de 30 de mayo del 2024 [7] al señalar que «son modelos de IA capaces de generar texto, imágenes u otros tipos de medios. Aprenden los patrones y la estructura de sus datos de entrenamiento de entrada y generan nuevos datos con características similares. Los avances en las redes neuronales profundas basadas en transformadores permiten que la IA generativa acepte como entrada indicaciones en lenguaje natural, incluidos los grandes modelos lingüísticos (LLM) como GPT-4, Gemini, Claude y LlaMA".

II.2.- BENEFICIOS DE LA IA GENERATIVA

La IA, en general, y la generativa en particular, se está integrando prácticamente en todos los ámbitos de la vida,[8] aportando

6 En el Informe se muestra la evolución de sus versiones, desde resolver tareas específicas a los modelos de base y los modelos fundacionales.

7 Nota 3 pág.3

8 La propia LIA lo expresa en su Considerando 4 al mostrar los beneficios de la IA: "La IA es un conjunto de tecnologías en rápida evolución que contribuye a generar beneficios económicos, medioambientales y sociales muy diversos en todos los sectores económicos y las actividades sociales. El uso de la IA puede proporcionar ventajas competitivas esenciales a las empresas y facilitar la obtención de resultados positivos desde el punto de vista social y medioambiental en los ámbitos de la asistencia sanitaria, la agricultura, la seguridad alimentaria, la educación y la formación, los medios de comunicación, el deporte, la cultura, la gestión de infraestructuras, la energía, el transporte y la logística, los servicios públicos, la seguridad, la justicia, la eficiencia de los recursos y la energía, el seguimiento ambiental, la conservación y restauración de la biodiversidad y los ecosistemas, y la mitigación del cambio climático y la adaptación a él, entre otros, al mejorar la predicción, optimizar las operaciones

múltiples beneficios, con un crecimiento exponencial[9]. La transformación de la IA generativa en la forma de generar contenidos va a afectar a todos los aspectos de nuestra forma de vivir, trabajar y jugar. Lo muestra el documento de debate Generative AI:Implications for Trust and Governance, 4, publicado en junio del 2023[10], que a su vez, reconoce que se ha descubierto el valor de los Chatbots como valiosos asistentes y el valor de la IA generativa en campañas de marketing de éxito, en el diseño de productos, de moda y de fármacos, en la mejora de los modelos de diagnóstico (sin afectar la privacidad de los pacientes), en la creación del gemelo del paciente, para aplicarse en medicina y precisión de ensayos clínicos. Muestra también su valor en la industria de entretenimiento, en el cine y en las plataformas de búsqueda en línea, con resultado que no serán hipervínculos sino conversacional. En el sector público, afirma el mencionado documento, hace a los ciudadanos más eficientes y accesibles los servicios públicos y puede ayudar en tareas de infraestructuras públicas y en el cambio climático de las ciudades inteligentes. Aporta beneficios como la posibilidad de mejorar la productividad, la accesibilidad y la diversidad de los contenidos mediáticos. Permite a su vez desarrollar nuevas creaciones o versiones mejoradas, agilizar procesos de ejecución, ofrecer nuevas líneas de negocio, etc.[11] La cultura se enriquece con sus creaciones.

y la asignación de los recursos, y personalizar las soluciones digitales que se encuentran a disposición de la población y las organizaciones".

9 Sobre el crecimiento de la misma véase, la encuesta e informe "El estado de la IA a principios de 2024: la adopción de la IA generativa aumenta y comienza a generar valor", McKinsey& Company, 30 de mayo de 2024, (SINCLA, A et al) (https://www.mckinsey.com/locations/south-america/latam/hispanoamerica-en-potencia/el-estado-de-la-ia-a-principios-de-2024-la-adopcion-de-la-ia-generativa-aumenta-y-comienza-a-generar-valor/es-CL) (recuperado el 2 de agosto de 2024).

10 Pág.7

11 Así afirman FRANGANILLO, J., "La inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos", methaodos.revista de ciencias sociales (2023) 11(2) m231102a1010.17502/mrcs.v11i2.710, pág.3. y SUÁREZ JAQUET, H. et HINOJAL CUADRADO, E. "El uso del deepfake en producciones audiovisuales: consideraciones jurídicas", coordinador: ORTEGA BURGOS, E., Propiedad intelectual, 2022, Documento TOL9.141.396, pág.1

Los propios artistas, a su vez, utilizan las redes neuronales para realizar sus obras.

En nuestro devenir diario aparece constantemente, por ejemplo, en buscadores y navegadores[12]. Son múltiples las apps que nos ofrecen sus beneficios (por ej. Chat GPT, Lensa, Dalle 2).

El Informe del SEPD "La IA generativa y el EUDPR. Primeras orientaciones del SEPD para garantizar el cumplimiento de la protección de datos al utilizar sistemas de IA", de 3 de junio de 2024, reconoce que "la IA generativa, al igual que otras tecnologías, ofrece soluciones en varios campos destinadas a apoyar y mejorar las capacidades humanas. Sin embargo, también plantean retos con posibles repercusiones en los derechos y libertades fundamentales que corren el riesgo de pasar desapercibidos, pasarse por alto o no ser debidamente considerados y evaluados"[13]. En este panorama de bondad, la IA generativa, conlleva también riesgos para los derechos fundamentales, la sociedad, la economía y la democracia.

II.3.- LOS RIESGOS DE LA IA GENERATIVA

Los contenidos generados por la IA generativa

Se van a distinguir diferentes riesgos provocados por la IA generativa[14]. Antes de analizar los puntos siguientes, hay que tener presente el concepto de datos personales, ya que pueden verse afectados por estas creaciones, como "toda información sobre una persona física identificada o identificable ("el interesado");

12 FRANGANILLO, J. "La inteligencia artificial...op. cit. pág.11 añade los paquetes ofimáticos, bases de datos científicas y programas de edición (imagen y vídeo).

13 Pág.6

14 Sobre esta materia en EEUU, véase el informe sobre el Marco de Gestión de Riesgos de la Inteligencia Artificial: Perfil de la Inteligencia Artificial Generativa, del NIST AI 600-I (borrador público inicial), abril 2024, NIST AI 600-I, julio de 2024 (recuperado el 2 de agosto del 2024).

se considerará persona física identificable toda persona cuya identidad pueda determinarse, directa o indirectamente, en particular mediante un identificador, como por ejemplo un nombre, un número de identificación, datos de localización, un identificador en línea o uno o varios elementos propios de la identidad física, fisiológica, genética, psíquica, económica, cultural o social de dicha persona" (art.4.1 Reglamento (UE) 2016/679 de Protección de Datos, concepto al que remite el art.3 definición 50 LIA). Conforme a esta definición, tal y como afirman SUÁREZ JAQUET, H et al. [15] la apariencia, la voz y los gestos entrarían en esta categoría de datos a efectos legales.

La IA generativa puede crear textos a través del uso de los modelos de lenguaje (por ejemplo Generative-Pre-trained Transformer y LLaMa), son tan reales que parecen auténticamente escritos por personas. Crea imágenes originales muy realistas, usando redes generativas antagónicas [16] con indicaciones en lenguaje natural.

La tecnología deepfake utiliza la IA generativa para crear vídeos sintéticos tan reales que pueden hacerte creer que lo irreal es real se califican así de hipertrucados. En la LIA se utiliza el término "ultrasuplantación" y se define como "un contenido de imagen, audio o vídeo generado o manipulado por una IA que se asemeja a personas, objetos, lugares, entidades o sucesos reales y que puede inducir a una persona a pensar erróneamente que son auténticos o verídicos" (concepto 60, art.3).

15 SUÁREZ JAQUET, H et al. "El uso…op.cit.pág.4. Sobre la evolución del derecho a la protección de datos de carácter personal: el algoritmo transparente y responsabilidad-accountability, véase BENDITO CAÑIZARES, M.T, "Estadio intermedio de reflexión para una futura regulación de la ética en el espacio digital europeo: los principios de transparencia y accountability", *Revista Aranzadi de Derecho y Nuevas Tecnologías* , núm 55/2021, BIB 2021/1465.

16 Sistemas de IA que se entrenan mediante Deep learning partiendo de gran cantidad de datos. Por ejemplo, el uso de Craiyon, DALL-E, Midjourney y Stable Diffusion, recogidos por MEIJOMIL, S., Inboundcycle, "6 generadores de imágenes con IA que no puedes perderte, 8 de noviembre del 2023, (https://www.inboundcycle.com/blog-de-inbound-marketing/generadores-de-imagenes-con-ia) (recuperado el 2 de mayo del 2024).

El término deepfake tiene su origen en la combinación de los términos deep learning y fake para referirse a un contenido falso generado con técnicas de aprendizaje profundo ("deep learning") [17]. Definido el deepfake por SUÁREZ JAQUET, H et al. Como "una técnica de IA que permite editar vídeos falsos de personas que aparentemente son reales, utilizando para ello algoritmos de aprendizaje no supervisados, conocidos en español como RGAs, y vídeos o imágenes ya existentes. El resultado final de dicha técnica es un vídeo muy realista, aunque ficticio"[18]. Los vídeos generados por esta técnica pueden "revivir" a personas fallecidas[19], hacer decir cosas que nunca se dijeron, hacer cosas que nunca hicieron o transformar totalmente la apariencia de los protagonistas. La manipulación de los vídeos, anteriormente se ceñía a trabajos manuales tediosos para modificar el contenido original (recortar, editar) pero no a las palabras o apariencia de los

17 El Estudio, la política europea frente a los deepfakes, de julio del 2021 lo recoge y señala su origen en Reddit (págs.2 y ss.). Lo muestra como un subconjunto de una categoría más amplia de «medios sintéticos» generados por IA, que no sólo incluye vídeo y audio, sino también fotos y texto. Sobre la diferencia entre deep fake y medios sintéticos, (págs.2 y ss.), Deepfake y tecnologías de medios sintéticos (págs.7 y ss): fotografía y videografía (pág.7), técnicas específicas de deepfake gráfico (págs.8 y ss.), de clonación de voz (págs.12 y 13) y síntesis de texto (13), nuevas tendencias y futuro y evolución de riesgos (págs.13 y ss.). Dicho Estudio desarrolla el panorama normativo europeo (y las lagunas) sobre los deepfakes (págs.37 y ss.) y las dimensiones de las medidas políticas para mitigar el impacto negativo de los deepfakes (dimensión tecnológica, de la creación, de la circulación, del objetivo y de la audiencia) (págs.58 y ss.).

18 SUÁREZ JAQUET, H et al. "El uso...op.cit.pág. 4 e 1 y 2. FRANGANILLO, J., "La inteligencia artificial...op.cit.pág.13, afirma que se nutre de datos e información derivada del comportamiento humano y, aplicada a determinados sectores y servicios, es capaz de imitarlo. El deepfake usa algoritmos de IA denominados Redes Generativas Antagónicas propios del "Deep learning. Añade en su pág.2 SUÁREZ JAQUET, H. et al. en qué consisten las RGAs. El Estudio, la política europea frente a los deepfakes, de julio del 2021, los define como medios sonoros o visuales manipulados o sintéticos que parecen auténticos, en los que aparece(n) una(s) persona(s) que parece(n) decir o hacer algo que nunca dijo (dijeron) o hizo (hicieron), producidos mediante inteligencia artificial o aprendizaje automático (págs.XIII y 2).

19 Por ejemplo, recientemente se ha vuelto a dar vida a Marilyn Monroe y Sean Connery en el thriller de espías, DUCK, de Rachel Maclean presentado en el Festival Internacional de Cine de Rotterdam 2024 y a Marilyn en solitario en un vídeo para ayudar a la gente a dormir. "Marilyn conoce a James Bond en la película Deepfake del artista", El informe de Marilyn, 5 de marzo de 2024, "https://themarilynreport-com.translate.goog/2024/03/05/marilyn-meets-james-bond-in-artists-deepfake-movie/?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=sc (recuperado el 2 de agosto del 2024).

"protagonistas"[20]. El poder recrear o revivir personajes de forma realista se originó en el cine con la técnica CGI (computer generated imagery) [21], muy utilizada además en publicidad; por ejemplo, "ha revivido" a Lola Flores, en la campaña publicitaria "Con mucho acento" (2021) de Cruzcampo por la que la agencia Ogilvy recibió el Gran Premio a la Eficacia y dos premios EFI de Oro.[22]

Por otra parte, la IA generativa puede crear vídeos sintéticos partiendo de unas simples instrucciones o indicaciones de texto, técnicamente menos evolucionados, sin perjuicio de una futura evolución más realista[23]. Otro contenido que genera esta IA es la voz. La voz es un elemento característico de la persona, como vimos, puede calificarse de dato personal, y por tanto el uso que se haga de ella por terceras personas debe ser responsable.

La síntesis de voz es una tecnología que permite convertir texto en una voz muy parecida a la humana que ya se estaba utilizando (asistentes de voz, voz en la navegación GPS...) aunque ha avanzado, ampliándose las posibilidades de su uso comercial (audiolibros, proyectos publicitarios, etc.), llegándose a utilizar voces sintéticas de personas fallecidas[24].

20 FRANGANILLO, J., "La inteligencia artificial...op.cit. pág.6

21 IBIDEM., pág.1, muestra cómo se sustituía digitalmente el rostro de un artista fallecido durante el rodaje por el de un doble y otra técnica: el "efecto Forrest Gump", originada en dicha película mezclando elementos históricos y actuales en las imágenes. SUÁREZ JAQUET, H et al. "El uso...op.cit.pág.2. muestran películas, en esa línea como, por ejemplo, Parque Jurásico y la Trilogía del Señor de los Anillos.

22 BORRACHERO GARRO, A. "Los retos de la tecnología en la publicidad: la campaña de Lola Flores", coordinador: ORTEGA BURGOS, E., *Propiedad intelectual*, 2022, Documento TOL9.141.406, págs.13 y ss., sobre la campaña de Lola Flores. Por ejemplo rejuvenece a personajes de ficción, con la tecnología digital rejuvenecedora ("de-aging technology") como el interpretado por Harrinson Ford en Indiana Jones y el dial del destino, y en películas como El curioso caso de Benjamin Button, también se utiliza para suplantar personas en programas satíricos de entretenimiento ("Entrevista por la cara"), hacer vídeos divertidos con rostros de estrellas (ej. Nicolas Cage) y suplanta a personajes de películas (FRANGANILLO, J., "La inteligencia artificial...op.cit.pág13 y SUÁREZ JAQUET, H et al. "El uso...op.cit.pág.3).

23 Como muestra FRANGANILLO, J., "La inteligencia artificial...op.cit. pág.10.

24 SUÁREZ JAQUET, H et al. "El uso...op.cit.pág.3, muestra como la Compañía: "Flawless AI" Software"TruySync" mejora el doblaje de las películas. Lo que también lleva

Desde el punto de las creaciones sintéticas musicales, puede tocar todos los puntos expuestos, ya que está por un lado la composición de letras de canciones, la música de las mismas, la posible utilización de voces sintéticas (de autores conocidos o no) y la realización de vídeos musicales[25]. Todo ello con los correspondientes problemas que pueden surgir en materia de propiedad intelectual. También se pueden crear, entre otros contenidos, códigos.

Los riesgos de desinformación y confusión

El que parezcan creaciones humanas los textos, imágenes[26], vídeos y voces originados por la IA generativa puede provocar, consciente o inconscientemente, desinformación, confusión, engaño, desconfianza y contenidos de mala o baja calidad, afectando, por ejemplo, al derecho a una información veraz[27].

posibles problemas legales ya que la interpretación está alterada al sustituir el movimiento de sus labios por los del actor de doblaje siendo necesario el consentimiento expreso del actor. Por otra parte, FRANGANILLO, J., "La inteligencia artificial...op.cit.pág.10, señala que se clonan voces de famosas para usarlas, con ética y con licencia, en audiolibros, productos audiovisuales y entornos inmersivos, voces en off o de doblaje (pionera la empresa Veritone que ofrece voces de celebridades, actores, actrices, deportistas y otras personas influyentes para cualquier proyecto sonoro o multimedia recibiendo royalties por el uso comercial de su voz). La usan medios de comunicación para otros fines por ej. para narrar la noticia con la voz de un periodista o presentador famoso.

25 DAVID, DemoCreator, "Los 10 mejores generadores de música IA gratis en 2024", 13 de marzo de 2024 (https://dc.wondershare.es/ai-voice/top-free-ai-music-generators.html) (recuperado el 7 de mayo del 2024). En ellas se puede dar un texto para que genere música, voces y vídeos musicales. En la lista que expone, señala (reconociendo además, en algunas, su posible comercialización) expresamente las que están libres de derechos de autor : Stability.ai, Beatoven.ai, Loudly, Soundful, Mubert. Añade otras sin aludir a los derechos de autor: Boomy, Soundraw.io, MusicaStar.AI., Riffusion, Suno AI. Posteriormente habla de la creación de vídeo musical con Wondershare Demoheator. Incide mucho en general,en la facilidad de uso, basta con unos clics.

26 La IA generativa se ha usado para crear "falsas" fotografías en prensa como ocurrió, entre otras, con las que se hicieron virales, entre marzo y abril del 2023, del expresidente Donald Trump forcejeando con la policía, tras la actuación de Midjourney otras de las escenas falsas eran del Papa Francisco con un abrigo de plumas o el abrazo de Yolanda Díaz y Pablo Iglesias. Esto provoca confusión, desconfianza, engaño y desinformación (Véase FRANGANILLO, J., "La inteligencia artificial...op.cit.pág.5 y SUÁREZ JAQUET, H et al. "El uso...op.cit.pág.2)

27 Sin embargo, la creación de imágenes conceptuales es menos problemática. FRANGANILLO, J., "La inteligencia artificial...op.cit.pág.12

Los deepfakes, sobre todo, generan "confusión" por el grado de sofisticación del falseamiento de la realidad en los vídeos. Se califica así este contenido de ultrafalso o hipertrucado. Reconoce FRANGANILLO que hoy la IA generativa es válida para situaciones que admiten cierto margen de error, cierta superficialidad argumental e incluso algún disparate, pero no lo es para cuestiones críticas (un trabajo científico, un consejo legal o financiero o una consulta médica). Produce "una engañosa ilusión de pensamiento racional", no entiende en un sentido humano nada de lo que escribe. Puede que su contenido sea incorrecto, ya que no dispone de un modelo de verdad y no siempre se apoya en fuentes fiables o evidencias robustas[28].

Un uso que, independientemente de los problemas legales que pueda generar, puede hacerse de forma maliciosa, por ejemplo, para abrir la puerta a los ciberdelitos a través de la suplantación de las personas. Debe tenerse presente como expone el Estudio, la política europea frente a los deepfakes, de julio del 2021 que estas actuaciones además de daños psicológicos y sociales conllevan daños financieros y perjuicios económicos (extorsión, robo de identidad, fraude, manipulación de precios de acciones, daños de marca y de reputación)[29].

Las creaciones sintéticas de IA generativa (imagen, vídeos, textos y voz) pueden como se ha visto, consciente o inconscientemente generar confusión y desinformación en el ciudadano al creer que son reales, por la perfección técnica alcanzada. Es difícil determinar su falsedad.

Son claros riesgos de estas tecnologías, como muestra BORRACHERO GARRO [30]por ejemplo, desde el punto de vista penal: la mencionada suplantación de personalidad (imagen y voz (deepvoice): para llamadas delictivas), enviar a la población mensajes

28 IBIDEM; pág.12

29 Págs.30 y 31.

30 BORRACHERO GARRO, A. "Los retos...op. cit. págs.17 y 18.

erróneos (fake news), imitar gestos o captar movimientos o señas de identidad con el fin de usarlos para dañar la reputación de una persona relevante (el deepfake es habitual en grabaciones pornográficas o campañas electorales con fines políticos). Es claro el uso de las mismas para generar contenido pornográfico, superponiendo la imagen de artistas famosas en un material pornográfico preexistente[31]. En el Estudio, la política europea frente a los deepfakes, de julio del 2021, se propone ampliar el marco jurídico actual en materia de delitos, se afirma que "teniendo en cuenta el daño que pueden causar los usos malintencionados de deepfakes, una evaluación de la solidez de las normas y reglamentos existentes a nivel de los Estados miembros podría ser útil para valorar si es necesario/ deseable añadir y especificar los delitos penales existentes. En Alemania, por ejemplo,

31 SUÁREZ JAQUET, H et al. "El uso…op.cit.pág.2. Como reitera FRANGANILLO, J. "La inteligencia artificial…op. cit. pág.9, "La amenaza es real: el 96%d de los vídeos deepfake publicados en línea en 2019 eran pornográficos y no consentidos, siendo las mujeres el colectivo más afectado (Ajder et al, 2019). Y en 2020 se identificaron más de 85.000 vídeos dañinos contra la reputación de figuras públicas, creados a un ritmo que se duplicaba cada seis meses (Ajder, 2020). Los algoritmos de aprendizaje profundo se entrenan con infinidad de imágenes que brinda internet, pero la tecnología avanza tan rápido que cada vez necesita menos datos de entrada para lograr un nivel similar de realismo (Giansiracusa, 2021)". Por su parte, VALERO, A., "Deepfakes Porn y violencia contra las mujeres", Fundación Cañada Blanch, 4 de junio de 2024, https://www.fundacioncanadablanch.org/noticias/deepfakes-porn-y-violencia-contra-las-mujeres/ (recuperado el 2 de agosto del 2024), muestra que el Informe State of Deepfakes 2023, de la empresa Home Security Heroes, afirma "que el 98% de los deepfakes que hay en Internet son pornográficos; que 7 de cada 10 sitios web pornográficos alojan deepfake porn y que El 99% de las personas que aparecen en las deepfake pornográficas son mujeres". En Almendralejo (Badajoz) unos menores compañeros de instituto o amigos, manipularon y difundieron imágenes, de un grupo de 20 chicas menores de edad, desnudas elaboradas con Inteligencia Artificial, (VIGARIO, D.,"Un año de libertad vigilada para los 15 jóvenes que manipularon y difundieron imágenes con IA de menores desnudas en Almendralejo", El mundo, 9 de julio de 2024 (recuperado en 1 de agosto del 2024)). Sobre este ejemplo y otros, en la misma línea en México, EEUU, véase LUCIO LÓPEZ, L.A., "Deep fake porn, la inteligencia artificial da nueva cara al ciberacoso escolar", Ciem, 2024. Recientemente aparecieron imágenes manipuladas con IA en X, deTaylor Swift desnuda (DURAN, I., "Taylor Swift y sus desnudos hechos con IA: CEO de Microsoft dice "hay que actuar ya" ante los deepfakes, Infobae, 29 de enero del 2024, https://www.infobae.com/tecno/2024/01/27/taylor-swift-y-sus-desnudos-hechos-ia-ceo-de-microsoft-dice-hay-que-actuar-ya-ante-los-deepfakes/ (rescatado el 3 de agosto del 2024)). Sobre este punto, ÁLVAREZ, P. y EGUILUZ, J. " El Reglamento de IA ante los deepfakes de desnudos", 2 de octubre del 2023, Cuatrecasas (https://www.cuatrecasas.com/es/spain/propiedad-intelectual/art/el-reglamento-de-ia-ante-los-deepfakes-de-desnudos) (recuperado el 1 de mayo del 2024).

está prohibida la distribución de un deepfake que viole los derechos de la persona (como la pornografía deepfake), pero no su producción"[32].

Riesgo de ataque a los derechos al honor y la propia imagen

De lo expuesto, se observa cómo se ven afectados, derechos fundamentales de la persona como el derecho al honor, la intimidad y la propia imagen (art.18 CE y LO 1/1982, De 5 de mayo, sobre protección al derecho al honor, a la intimidad personal y a la propia imagen). Nuestra privacidad, nuestros datos personales se ven atacados[33] para utilizarlos, de forma maliciosa o no, incluso delictiva. Ponen en tu boca palabras que nunca dijiste y tu imagen realiza hechos que nunca hiciste, creando falsas o perversas realidades, tengas o no una proyección pública (periodista, político, famosos...) afecta a tu reputación y credibilidad. Pueden ser el hilo transmisor de noticias falsas manipulando la opinión pública, provocando la desinformación.

Tenemos un derecho personalísimo a la propia imagen, que incluye la apariencia física, la voz, o la semejanza o parecido físico, (en definitiva "datos personales"[34]), además del derecho al honor y la intimidad. Datos que se están utilizando. La única forma de admitir el uso por terceros de estos datos personales es con el necesario consentimiento de los dueños de la imagen o la voz, o de sus herederos para que no se produzca una intromisión ilegítima al derecho al honor, la intimidad y la propia imagen[35]. El

32 Estudio, la política europea frente a los deepfakes. Panel para el futuro de la Ciencia y la Tecnología dirigido por Philip Boucher, EPRS/ Servicio de Estudios del Parlamento Europeo, Unidad de Prospectiva Científica (STOA), PE 690.039-julio de 2021, pág.61. Seguido por ÁLVAREZ, P. y EGUILUZ, J. "El Reglamento de IA ante...op.cit.

33 SUÁREZ JAQUET, H. et al. "El uso...op. cit. pág.1

34 IBIDEM; pág.5.

35 Con respecto al derecho fundamental a la propia imagen como expresa BORRACHERO GARRO, A. ("Los retos...op. cit. págs.15 y 16): Se produce un desdoblamiento del derecho fundamental a la propia imagen. El derecho fundamental a la propia imagen

uso de la voz, imágenes y vídeos de personas fallecidas genera además un debate ético, aunque jurídicamente se puede legalizar dicho uso si lo consienten los herederos, pero ¿hasta qué punto lo habrían querido ellas?[36]

Por otra parte, es necesario que expresamente se identifique cuando una IA generativa es la autora de la creación, siendo una creación sintética y no humana. Se evita así la confusión, como ha sucedido, por ejemplo, en concursos de arte y fotografía, en los que se desconocía el origen sintético de las creaciones premiadas[37](por ejemplo, en materia de doblaje se pide que se distinga al oírla - que tenga "acento de IA"-, otra forma sería insertar una huella digital indeleble indicándolo). Es esencial también,

---

se extingue con la muerte pero subsiste la protección a la memoria del fallecido (su tutela corresponde a las personas que establece el art.4.2 LO 1/1982)(SUÁREZ JAQUET, H et al. "El uso…op.cit.pág.5.). El aspecto patrimonial del derecho de imagen, es un bien jurídico diferente que no está incluido en el contenido esencial del derecho fundamental y puede protegerse a nivel de legalidad ordinaria (aunque dicha Ley no contempla expresamente la transmisión patrimonial mortis causa puede tener cabida en nuestro ordenamiento jurídico), para evitar conflictos lo mejor es que "se recabe el consentimiento de los titulares del derecho de imagen de la persona fallecida refiriéndonos a la vertiente patrimonial a la vez que se solicita una renuncia adicional a emprender acciones legales por la defensa del derecho fundamental". Muestran SUÁREZ JAQUET, H et al. ("El uso…op.cit.págs. 4 y 5) como se considera "intromisión ilegítima" la "captación , reproducción o publicación por fotografía, filme o cualquier otro procedimiento, de la imagen de una persona en lugares o momentos de su vida privada o fuera de ellos" (art.7.5 LO 1/1982) y el art.7.6 LO 1/1982 extiende la consideración de intromisión ilegítima" a la utilización del nombre, de la voz o de la imagen de una persona para fines publicitarios, comerciales o de naturaleza análoga". Expresan que el derecho a la propia imagen no es absoluto, puede entrar en conflicto con la libertad de expresión del creador (constitucional también), que protege al mismo para difundir sus ideas, pensamientos y opiniones de forma libre y sin censuras, pero en el caso del deepfake no vale ya que hay "riesgo de confusión", imagen irreal pero que quiere que se perciba como real. Sí se considera, por ejemplo, legítima según el art. 8.2 b) LO/1982" la utilización de la caricatura de personajes públicos, de acuerdo con el uso social". El derecho a la propia imagen y al honor prevalece sobre el de libertad de expresión cuando hay descrédito o trato vejatorio de la persona.

36 Véase BORRACHERO GARRO, A. "Los retos…op.cit. pág.19; FRANGANILLO, J., "La inteligencia artificial…op.cit.pág.13.

37 Por ejemplo, el premio Sony Wordl Photography Awards, lo ganó Boris Eldagsen, lo rechazo posteriormente al reconocer que no era una fotografía (MORAN, I., Photolari, 2 de mayo, 2024, "20.000 euros por la imagen IA que engañó al jurado de los Sony Wordl Photography Awards el año pasado" ((https://www.photolari.com/20-000-euros-por-la-imagen-ia-que-engano-al-jurado-de-los-sony-world-photography-awards-el-ano-pasado/) (recuperado el 20 de mayo del 2024).

establecer mecanismos de verificación y transparencia que permitan identificar la fuente y la autenticidad de las grabaciones sonoras[38]. No debemos olvidar que cualquiera actualmente puede acceder a esta IA generativa fácilmente, sin conocimientos y sin gasto o con un pequeño gasto (lo cierto es que esto favorece la democratización cuando su uso es correcto). Se requiere un uso responsable.

Las personas tienden a sobreestimar su capacidad de detectar engaños o manipulaciones, lo que les lleva a carecer de una actitud crítica. Por ejemplo, se incrementan las fake news a las que el público les concede absoluta credibilidad[39]. Por ello, además de la formación del público ante estas situaciones, la ingeniería debe intensificar los avances técnicos, la creación de herramientas proactivas de detección y buscar modos de informarle de que el contenido es sintético y potencialmente malicioso[40].

El estudio, la política europea frente a los deepfakes, de julio del 2021[41] destaca entre sus preocupaciones, la detección (manual o automática) y la prevención de deepfakes: "Debido al papel central que desempeñan las plataformas en línea y otros intermediarios en la difusión de deepfakes, la Comisión plantea obligar a dichas plataformas e intermediarios a disponer de un software de detección de deepfakes como requisito previo para un posible

38 FRANGANILLO, J. "La inteligencia artificial...op. cit. pág.13.

39 IBIDEM., pág..13. Véase Documento de debate Generative AI:Implications for Trust and Governance, 4, publicado en junio del 2023, pág.16

40 FRANGANILLO, J. "La inteligencia artificial...op. cit. pág.13

41 En el Estudio, la política europea frente a los deepfakes de julio de 2021, sobre la detección (manual o automática) y prevención (ataques contra los algoritmos de deepfake, refuerzo de marcadores de autenticidad de los materiales audiovisuales y ayudas técnicas para que la gente los detecte más facilmente) de deep fakes se resalta que no es difícil evadir la detección (p. III y ss, y 17 y ss.). Sobre los sistemas de "marcado" ESPUGA TORNÉ, G. "Cómo identificar contenido generado por IA" Linkedin, junio, 2024 (recuperado el 10 de julio). Herramientas que permiten detectar textos generados por Inteligencia Artificial (Metadatos, Marcas de agua, Fingerprinting, herramientas de detección). HOLCOMBE, J., "Las 9 Mejores Herramientas de Detección de Contenidos con IA que tienes que conocer", Kinsta, 5 de abril del 2023 https://kinsta.com/es/blog/deteccion-de-contenidos-ia/ (recuperada el 6 de julio del 2024).

etiquetado. Una alternativa para la detección es el uso de filtros de carga –por ejemplo, que los rostros aparezcan difuminados hasta que las personas retratadas den su consentimiento–. Aunque esta opción también conlleva inconvenientes como los límites de la técnica o los peligros de censurar expresiones artísticas, vulnerando la libertad de expresión"[42]. Se había planteado, la futura regulación de las IA generativas en la reunión del G7 de 19-21 de mayo de 2023[43], viendo los riesgos de las mismas y ante la posibilidad de que la falta de regulación permitiera que las grandes compañías empezaran a autorregular el entrenamiento de sus IA generativas (sin olvidar la participación Italiana que prohibió temporalmente el uso de Chat GPT).

No obstante, la tecnología deepfake también se ha utilizado para concienciar al público de sus riesgos, que es necesario contrastar la información que aparece en internet y en los medios de comunicación. Ser diligentes e identificar las fuentes fiables de las que no lo son tal y como afirman SUÁREZ JAQUET, H et al [44].

42 Pág.62.

43 GARAY, J. ,Wired, 19 de mayo del 2023,"El G 7 promete regular las IA generativas antes de que termine el 2023"(https://es.wired.com/articulos/g7-promete-regular-las-ia-generativas-antes-de-que-termine-el-2023) (recuperado el 30 de mayo del 2024) , muestra esos efectos negativos: Los delitos que la utilizan y las consecuencias sociales, son evidentes y el mal uso de las mismas: por ejemplo deepfakes sexuales no consensuados, estafas que usan la voz de los familiares de las víctimas, imágenes que nunca existieron que provocan desinformación y la filtración de conversaciones privadas con chatbots. Exponen los miembros del G7 que debe regularse la IA y reconocen los riesgos y problemas que hay con las IA generativas. Las naciones más avanzadas del mundo no esperarán para que las grandes compañías (Open AI, Meta (LlaMA) Y Google (PaLM) establezcan reglas para el entrenamiento de sus IA generativas. Se establece "el proceso de la IA de Hiroshima".

44 SUÁREZ JAQUET, H et al. "El uso...op.cit.pág.2, el ejemplo al que se refieren es el del cómico Jordan Peele que en 2018 creó un vídeo en el que aparecía el expresidente de los Estados Unidos Barack Obama señalando cómo esta tecnología nos podía hacer creer cosas que nunca diríamos, al tiempo que estaba insultando a Donald Trump.

Riesgo de infracción o vulneración de los derechos de autor y la moderación de contenidos

La IA generativa, crea contenidos partiendo de datos ya existentes (entrenando con ellos). Para entrenar los sistemas de aprendizaje profundo se utiliza el raspado de datos web scraping, se recopila automáticamente un volumen grande de información accesible públicamente. Se cuestiona la licitud y legalidad de nutrirse con obras que normalmente tienen derecho de autor. Los propios creadores han contribuido al sistema al dar acceso a través de internet a sus obras. En la Unión Europea, el Reglamento General de Protección de Datos no establece que ese método sea ilegal, si bien restringe lo que se puede hacer con los datos extraídos. Una minería de textos avalada legalmente cuando su finalidad es la investigación científica no comercial[45].

Determinar la originalidad de las creaciones de la IA generativa, es complicado, no sabemos a ciencia cierta cuando se violan los derechos de autor de obras anteriores, por plagio. Otro problema es determinar quién es el "creador", el copyright, como señala MARTÍNEZ ESPÍN[46], el autor se plantea si las creaciones generadas por sistemas de Inteligencia Artificial son obras protegidas por derechos de autor. Afirma que el derecho de autor protege cualquier creación de la mente. Pero ¿mente humana o máquina? Preguntándose a su vez el autor si la creación de un sistema de IA ¿es una obra? ¿quién es el autor? ¿el sistema de IA, el programador o el usuario?

45 FRANGANILLO, J. "La inteligencia artificial...op. cit. pág.5. Sobre la minería de datos y textos véase MUÑOZ VELA, J. M, "Inteligencia artificial generativa. Desafíos para la propiedad intelectual", *Revista de Derecho UNED*, núm.33, 2024, Premio de artículos jurídicos "García Goyena", 22ª convocatoria (curso 2022-2023), Facultad de Derecho. UNED, págs.35 y ss.

46 MARTÍNEZ ESPÍN, P.,La propuesta de marco regulador de los sistemas de inteligencia artificial en el mercado de la UE ", Revista CESCO de Derecho de Consumo, nº46/2023, pág.3, (doi.org/10.18239/RDC_2023.46.3322).

En el mundo de la música, se han producido graves problemas desde el punto de la originalidad de las canciones creadas a través de la IA generativa, no solo por la utilización de parte de canciones existentes o música anterior para poder crearla, sino también, por la utilización de voz sintética de algunos autores sin que ellos lo sepan y hayan consentido.

FRANGANILLO[47] expresa que se podría proteger la propiedad intelectual señalando explícitamente la IA las fuentes que ha utilizado y que han entrenado al sistema. Manifiesta el autor que esta situación reclama una regulación y establecer formas de compensación para garantizar la viabilidad de la industria de medios y el ecosistema cultural del que se aprovechan hoy gratuitamente los modelos generativos. Las empresas de síntesis de voz han creado normas éticas y se aseguran de que la persona preste el consentimiento para que se clone su voz, controlando su uso y recibiendo una retribución. Debe, así, contarse con el consentimiento del dueño de la voz (o la imagen y resto de datos personales). Recientemente la actriz Scarlett Johanson paralizó el uso no consentido de una voz que "replicaba con gran exactitud" la suya cuando previamente había rechazado la propuesta de que se utilizara su voz en la versión más avanzada de ChatGPT (ChatGPT 40)[48].

MUÑOZ VELA[49] reconoce que los sistemas inteligentes generativos tienen capacidad creadora pero no creativa y que la IA actual carece de autonomía efectiva. Manifiesta, que dichos sistemas

47 FRANGANILLO, J. "La inteligencia artificial...op. cit. pág.15

48 MCMAHON, L., BBC News, 20 de mayo de 2024, "Cuando la escuché me quedé en shock": por qué el programa de IA ChatGPT dejará de usar la voz que se parece a la de Scarlett Johanson, https://www.bbc.com/mundo/articles/cprrzn8g2wqo.) (recuperado el 30 de mayo del 2024) en el artículo se muestra como inicialmente se le hizo una oferta para que prestara su voz a la nueva versión de Chat GPT. 40, se negó, y se utilizó una voz tan similar a la suya, la voz "Sky", teniendo en cuenta, el avance de las funciones de voz, su avance conversacional de Chat GPT 40,que la actriz ha conseguido la paralización de su uso.

49 MUÑOZ VELA, J. M, "Inteligencia artificial generativa. Desafíos...op.cit. págs.65 y ss., pág.67.

no pueden garantizar en el momento presente la originalidad y singularidad de sus resultados. Añade, que los marcos reguladores de la propiedad intelectual requieren la autoría humana y la originalidad del resultado para su protección (derechos de autor y conexos), negándola a las creaciones absolutamente artificiales, sin intervención humana o con intervención no relevante (en estos casos la IA no tiene la condición de autor ni su protección). Se plantea, si estos resultados tendrían protección a través del derecho de autor o del de la propiedad u otros. Expresa el autor que los datos y contenidos de los que se nutren los sistemas inteligentes (inputs), sí están protegidos, su uso requiere la autorización expresa del titular (salvo excepciones o limitaciones legalmente establecidas). Reconoce también la protección a través del derecho de autor de los prompts suministrados al sistema para crear el resultado (ouput), "si la secuencia de instrucciones evidencia un esfuerzo y complejidad cualitativa y cuantitativa que pueda determinar un resultado único y a las que se pueda asociar la autoría humana y la originalidad, considero que podrían resultar protegibles como PI, incluso como obra literaria/técnica desde su creación".

El Informe del SEPD 2024, en relación al consentimiento expresa que "El tratamiento de datos personales en el contexto de los sistemas de IA generativa requiere una base jurídica acorde con el Reglamento. Si el tratamiento de datos se basa en una obligación legal o en el ejercicio de la autoridad pública, dicha base jurídica debe establecerse de forma clara y precisa en la legislación de la UE. El uso del consentimiento como base jurídica requiere una cuidadosa consideración para garantizar que cumple los requisitos del Reglamento, a fin de que sea válido"[50].

50 Pág.17.En el Considerando 105 LIA se muestra que cuando se utilizan en técnicas de prospección de textos y datos, contenidos protegidos por el derecho de autor, se requiere la autorización del titular, "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 en determinadas circunstancias".

Debe tenerse presente la responsabilidad de los prestadores de servicios, como establecen SUÁREZ JAQUET, H et al.[51], en virtud del art.73 del Real Decreto-ley 24/2021, de 2 de noviembre, se transpone al ordenamiento español las Directivas sobre derechos de autor y derechos afines en el mercado digital, en el uso de contenidos protegidos por parte de prestadores de servicios para compartir contenidos en línea (Instagram, Facebook, Twitter). Ellos tienen que obtener las correspondientes autorizaciones de los titulares de los derechos, si no se las otorgan responderán de los actos no autorizados de comunicación al público, salvo que demostraran que hicieron sus mayores esfuerzos por obtenerla y para garantizar la indisponibilidad de la obra y prestaciones. Por ello, como afirman los autores, si un deepfak no ha obtenido las autorizaciones pertinentes y esta circunstancia se verifica por los prestadores de servicios, podría considerarse como contenido que vulnera los derechos de autor y se debería retirar de la plataforma, para no asumir responsabilidad.

Recientemente Meta, en Instagram, por ejemplo, ha lanzado un formulario a sus usuarios para que expresen que no consienten que se utilicen sus datos (que están en la aplicación), para el entrenamiento de la IA, pero no ha comunicado expresamente esta opción. Dada la transcendencia, debería haberse informado de una manera más directa e individual, en la que quede un conocimiento claro de lo que implica aportar dichos datos personales al entrenamiento de la IA. Imágenes, vídeos, etc. subidos a la plataforma que pueden usarse y aparecer de una forma u otra donde menos te lo esperas. Siendo una autorización consciente y necesaria. Tendría que ser requisito de acceso a la plataforma el responder a dicha solicitud, se forzaría así obtener el consentimiento o la negativa de forma segura. Esta actuación se ha parado finalmente por Meta. La autoridad de protección de datos de Irlanda, Data Protection Commission

51 SUÁREZ JAQUET, H et al. "El uso...op.cit.pág.7

(DPC) en nombre de la Unión Europea no la ha permitido al afectar a la politica de privacidad. Por otra parte, la Fiscalía española abrió recientemente diligencias por esta actuación que ya había cesado. La actuación y normativa europea ha llevado a Meta a afirmar que no ofrecerá sus nuevos modelos de IA generativa en Europa[52].

GOLDMAN SACHS [53], destaca otra problemática de la IA generativa conectada con este punto, la moderación de contenidos. Muestra como las plataformas de redes sociales y sitios web, donde se publican contenidos por los usuarios no se hacen legalmente responsables de ellos (estableciendo una especie de escudo legal) y sí a los usuarios.

Riesgo del mercado laboral

La IA generativa sigue la estela de la IA en general, que transformará, en todos los sentidos, el ámbito laboral (con la creación, cualificación y desaparición de empleos. Nadie "tiene segura la silla", ni los CEOs[54]), e incluso afectará a los sistemas de seguridad

52 Sobre este punto, RAA J., "Meta detiene su proyecto para entrenar a la IA con publicaciones de Facebook e Instagram en Europa", Tecnología, El País, 14 de junio de 2024 (recuperado el 1 de julio del 2024) ; AFP, "La fiscalía investiga si Meta vulnera la protección de datos de sus usuarios", El Mundo, Empresas, 4 de julio de 2024 (recuperado el 7 de julio del 2024) y PASCUAL, M.G.,Meta no ofrecerá sus nuevos modelos de IA generativa en Europa por su "impredicible entorno regulatorio", El Pais, Tecnología, 18 de julio de 2024 (recuperado el 19 de julio del 2024).

53 GOLDMAN SACHS, Principales riesgos que entraña la inteligencia artificial generativa: enumeración, fundspeople, (https://fundspeople.com/es/principales-riesgos-que-entrana-la-inteligencia-artificial-generativa/), 26 de abril de 2023 (rescatada en 20 de abril del 2024). Goldman Sachs ha publicado un white titulado Generative AI- Part I: Laying Out the investment Framewor, en dicho trabajo académico los autores analizan los riesgos asociados a la IA generativa, rama que se centra en la generación de contenido original a partir de datos existentes. (5 riesgos: moderación de contenidos, desinformación, infracción de los derechos de autor, privacidad, cuestiones éticas).

54 Los chatbots están siendo utilizados para tomar decisiones de directivos que podrían ver peligrar sus puestos de trabajo, VIDAL. M, en LinkedIn recoge una noticia, el New York Times explora la tendencia de incorporar la IA como directora de algunas empresas, reconoce el potencial de la IA para reemplazar a los CEOs. (Fuente: https://www.nytimes.com/2024/05/28/technology/ai-chief-executives.html) (recuperado el 7 de junio de 2024).

social. Los gobiernos deberán ayudar a capacitar digitalmente a los trabajadores[55].

La utilización de la IA por su automatización y aumento de productividad puede hacer innecesarios ciertos trabajos, destruyendo así empleos. En materia cultural, de medios de comunicación y audiovisual, como muestra FRANGANILLO, están especialmente en peligro los diseñadores, los ilustradores, los fotógrafos, los dobladores y los guionistas [56]. En estos casos, nos encontramos con tecnología al alcance de cualquiera sin conocimiento de diseño gráfico ni de otro tipo al efecto (fotografía, etc.,). Puede crear contenido sintético (aun sin vulnerar la imagen de una persona), afectando al trabajo de los mencionados profesionales, que han visto devaluada su habilidad para presentar trabajos de calidad[57]. Podemos añadir que la creación de códigos con la IA generativa puede afectar al trabajo de los creadores de páginas web y a los programadores.

El mencionado autor muestra como la comercialización de estas creaciones generan ganancias y que, por ello, algunos bancos de imágenes han reescrito sus directrices para impedir la venta de materiales creados por la IA generativa. Termina reconociendo que la IA, para entrenar sus modelos, se abastece de trabajos

También para preparar las entrevistas de trabajo (ANDRÉS, R., "La última tendencia para bordar las entrevistas de trabajo: entrenar con ChatGPT como reclutador", Xataca, 6 de marzo del 2024 (recuperado el 28 de julio del 2024)).

55 Un estudio detallado de la evolución de la IA en el ámbito laboral lo ofrece MUÑOZ VELA, J.M, *Retos, riesgos, responsabilidad y regulación de la inteligencia artificial. Un enfoque de seguridad física, lógica, moral y jurídica*, Thomson Reuters- Aranzadi, Pamplona, 2022, págs.36 y ss., aportando las teorías que ven una mejora con la misma y otras más catastrofistas. Por otra parte, La LIA recoge la importancia del empleo en los Considerandos 2,9,57 y 58 y en el Anexo III apartado 4.

56 FRANGANILLO, J. "La inteligencia artificial...op. cit. págs.13 y 17, recoge también la huelga del Sindicato de guionistas de Estados Unidos, que exige, entre otras demandas, que la IA no sustituya labores creativas, ni escriba ni reescriba material literario, ni se entrene con obras de guionistas y que el 36% de los trabajadores de la industria estadounidense del entretenimiento temen el impacto de la IA generativa en sus empleos, sobre todo por la vulneración de la propiedad intelectual (pág.17).

57 Véase IBIDEM.,. pág.17.

ajenos, afectando al derecho a la propiedad intelectual. Un "deber legal ético" no resuelto que exigiría recibir una compensación por su uso. El problema surge si sustituye el servicio de los profesionales. La Asociación de Medios de Información exige una nueva tasa por el aprovechamiento que la IA hace de sus contenidos sin reconocimiento ni retribución[58].

Riesgo para el medio ambiente

El derecho al medio ambiente es esencial. Se afirma en el Considerando 48 LIA, que hay que tener en cuenta, cuando se evalúe la gravedad del perjuicio que puede ocasionar un sistema de IA, el derecho fundamental a un nivel elevado de protección del medio ambiente consagrado en la Carta y aplicado en las políticas de la Unión.

La IA generativa, al igual que el resto de IAs y de tecnologías como Blockchain, tienen un gran consumo energético.

Nos muestra FIGUEROA como la IA generativa en sus creaciones consume mucha energía y contamina, al emitir gran cantidad de carbono. Por ejemplo, la acción que más consume es la creación de imágenes, comparada con la de texto[59], una imagen equivale a la carga de un celular. Lo afirma, como muestra FIGUEROA, un Estudio del startup IA Hugging Face y la Universidad Carnegie Mellon de Estados Unidos, dirigido por Sasha Luccioni. Descubrió

58 DEL CASTILLO, C., "Los creadores del canon AEDE quieren una "tasa ChatGPT" para la inteligencia artificial, el Diario.es, 3 de mayo del 2023, (https://www.eldiario.es/tecnologia/creadores-canon-aede-quieren-tasa-chatgpt-inteligencia-artificial_1_10171676.html), (recuperado el 20 de mayo del 2024).

59 Siguiendo a FIGUEROA, J.C., "Crear una sola imagen con inteligencia artificial consume tanta energía como cargar tu teléfono", Hipertextual, Tecnología, 12 de diciembre de 2023 (https://hipertextual.com/2023/12/crear-imagen-con-inteligencia-artificial-consume-esta-energia) (recuperado el 20 de abril del 2024), el autor muestra que el estudio de la startup IA Hugging Face y la Universidad Carnegie Mellon de Estados Unidos, afirma que: ejecutar 1.000 acciones de generación de texto, en una herramienta como ChatGPT, solo consume tanta energía como el 16% de la batería de tu celular. El autor a lo largo del artículo ofrece datos más detallados.

el estudio, a su vez, que el uso de grandes modelos generativos, como ChatGPT o Bard, consumía mucha más energía que los modelos de inteligencia artificial más pequeños diseñados para tareas específicas. El motivo es porque intentan hacer muchas cosas a la vez: generar, clasificar y resumir texto, en lugar de una sola tarea. Además, sus emisiones diarias exceden con creces las generadas durante el entrenamiento de modelos grandes. El impacto ambiental ofrece distintos frentes, por ejemplo, para enfriar los servidores de productos con ChatGPT son impresionantes las cantidades de agua necesarias. La industria busca cómo mitigar el impacto (principalmente por motivos económicos para reducir gastos, aunque conllevará una mejora en la calidad de vida) desarrollando estrategias, concretamente a través de la energía nuclear[60].

Riesgo por los sesgos (la necesaria alfabetización y democratización)

Otro de los riesgos a los que nos expone la IA generativa son los sesgos. La LIA no recoge la definición de sesgo, pero si hace alusiones expresas a los mismos[61].

Como expone FRANGANILLO[62] en conexión con la IA generativa (enfocada por él principalmente en el mundo audiovisual), los algoritmos de generación de imágenes, entrenados con datos sin filtrar, pueden reproducir estereotipos raciales (étnicos), culturales y de género incluidos en los datos, provocando sesgos. Los algoritmos no son neutrales y conviene corregir sus desviaciones para garantizar el principio de justicia y una actuación ética.

60 IBIDEM., recoge que Microsoft en mayo llegó a un acuerdo de compra de energía con Helion, un startup de fusión nuclear, para comprarle electricidad a partir de 2028 y también Sam Altman, director ejecutivo de OpenAI, quien también es uno de los inversores más importantes en Helion.

61 Concretamente en los Considerandos 27, 61,67, 70,110, arts.10.2 f) y g) y 5 a), e) y f); 14.4.b), 70.1, y Anexo XI, sección primera 2.c) LIA.

62 FRANGANILLO, J." La inteligencia artificial...op. cit. pág.13.

El sistema se debe calibrar y reentrenar. El autor muestra otros dos sesgos: un sesgo cognitivo, al tenerse la percepción pública distorsionada (por el cine y la literatura) por la creencia de que lo que dice la IA es cierto y otro sesgo porque la IA generativa suele seleccionar de la información accesible en internet la de fuente inglesa considerándola de mayor calidad, frente a la diversidad cultural. Se está intentando solucionar complementando las respuestas con el resultado de búsqueda en tiempo real.

Reconoce GOLDMAN SACHS [63], a su vez, que la precisión se puede mejorar si vuelven a entrenar con información mejorada pero el sesgo es más difícil al ser los seres humanos lo que entrenan estos modelos de IA. Un mal entrenamiento de los datos puede originar datos inexactos y sesgados.

Sobre los sesgos, el Informe del SEPD 2024, manifiesta el carácter prioritario de su minimización, al afirmar que "La aplicación de procedimientos y mejoras prácticas para minimizar y mitigar los sesgos debería ser una prioridad en todas las fases del ciclo de vida de los sistemas generativos de IA, para garantizar un procesamiento justo y evitar prácticas discriminatorias. Para ello, es necesario supervisar y comprender cómo funcionan los algoritmos y los datos utilizados para entrenar el modelo"[64]. Se habla de minimización y mitigación ya que son actualmente imposibles de erradicar de forma absoluta por la intervención del ser humano.

En la LIA se ve la preocupación por los sesgos que puedan afectar a la salud, la seguridad de las personas y negativamente a los derechos fundamentales o dar lugar a algún tipo de discriminación prohibida por el Derecho de la Unión. Especialmente ocurre cuando las salidas de datos influyan en las informaciones de entrada de futuras operaciones (art.10.2 f) LIA). Alude la LIA

63 GOLDMAN SACHS, "Principales...op.cit.

64 Pág.31

a las medidas adecuadas para detectarlos y prevenirlos[65] en los arts.10.2 g) y 10.5 a), e) y f) en conexión con los sistemas de IA de alto riesgo. Se muestra en el art.14.4 b) LIA un tipo de sesgo "el sesgo de automatización": "la tendencia a confiar automáticamente o en exceso en los resultados de salida generados por un sistema de IA de alto riesgo" y se pide ser conscientes del mismo. También se conectan los sesgos con los sistemas generales con riesgos sistémicos[66].

Los sesgos nos llevan a la manipulación y la desinformación de nuevo[67].

El uso correcto de la IA generativa pasa por entenderla, comprenderla por todos los agentes de la cadena que entran en contacto con la misma. Por ello, los gobiernos deben proporcionar programas de concienciación para que el público comprenda el alcance de la IA generativa y aprenda a defenderse de las nuevas falsificaciones. La educación es esencial para preparar la ciudadanía ante la proliferación de contenidos artificiales, aprendiendo sobre la IA y capacitando a la sociedad para adaptarse a sus efectos[68]. La capacitación en el conocimiento de la IA generativa puede además facilitar la apertura a la nueva forma de concebir el empleo que está aportando la IA.

En la reciente Convención Marco del Consejo de Europa sobre Inteligencia Artificial y Derechos Humanos, Democracia y Estado de Derecho de 17 de mayo del 2024[69] , su art.20 sobre alfabetización y competencias digitales, promueve ambas, dispone que "cada Parte fomentará y promoverá la alfabetización digital

65 Véanse los Considerandos 67 y 70 de LIA.

66 El Considerando 110 LIA, en relación a los sistemas generales con riesgos sistémicos reconoce que los tienen: "los modelos pueden dar lugar a sesgos dañinos y discriminación que entrañan riesgos para las personas".

67 Sobre los sesgos véase MUÑOZ VELA, J.M, *Retos*...op.cit.págs.56 y ss.

68 FRANGANILLO, J. "La inteligencia artificial...op. cit. pág.15.

69 13ª Sesión del Comité de Ministros (Estrasburgo, 17 de mayo de 2024), Comisión de Inteligencia Artificial (CAI), CM (2024)52-final.

y las competencias digitales adecuadas para todos los segmentos de la población, incluidas las competencias especializadas específicas para los responsables de la identificación, evaluación, prevención y mitigación de los riesgos planteados por los sistemas de inteligencia artificial".

La LIA muestra la necesaria alfabetización. Recoge expresamente que debe dotar a los proveedores, responsables del despliegue y personas afectadas de los conceptos necesarios para tomar decisiones con conocimiento de causa en relación con los sistemas de IA (Considerando 20).[70]Se define en el art.3 número 56 la "alfabetización en materia de IA" como "las capacidades, los conocimientos y la comprensión que permiten a los proveedores, responsables del despliegue y demás personas afectadas, teniendo en cuenta sus respectivos derechos y obligaciones en el contexto del presente Reglamento, llevar a cabo un despliegue informado de los sistemas de IA y tomar conciencia de las oportunidades y los riesgos que plantea la IA, así como de los perjuicios que puede causar" y el art.4 establece sobre la alfabetización en materia de IA que "Los proveedores y responsables del despliegue de sistemas de IA adoptarán medidas para garantizar que, en la mayor medida posible, su personal y demás personas que se encarguen en su nombre del funcionamiento y la utilización de sistemas de IA tengan un nivel suficiente de alfabetización en materia de IA, teniendo en cuenta sus conocimientos técnicos, su experiencia, su educación y su formación, así como el contexto previsto de uso de los sistemas de IA y las personas o los colectivos de personas en que se van a utilizar dichos sistemas".

70 Expone el Considerando 20 que "Con el fin de obtener los mayores beneficios de los sistemas de IA, protegiendo al mismo tiempo los derechos fundamentales, la salud y la seguridad, y de posibilitar el control democrático, la alfabetización en materia de IA debe dotar a los proveedores, responsables del despliegue y personas afectadas de los conceptos necesarios para tomar decisiones con conocimiento de causa en relación con los sistemas de IA" desarrolla posteriormente los conceptos y la actuación en el Considerando, conceptos que afectan a todos los agentes pertinentes de la cadena de valor de la IA.

Existe una conexión con la democratización, porque, aunque la IA está llegando rápidamente a todos los públicos, no se puede afirmar que se esté democratizando totalmente, ya que esos públicos no sólo han de poder usarla, sino que también deben poder entenderla y por eso aboga por una educación digital de conocimiento y de pensamiento crítico[71]. La democratización implica velar tanto por el acceso como por la educación, la ética y la transparencia. Hay que impedir que los intereses privados perjudiquen este proceso[72].

El riesgo de uso malicioso

La naturaleza humana puede dar a las nuevas tecnologías un uso reprobable, imprudente o malicioso. Por ello, las organizaciones que desarrollan aplicaciones de ella deben ser transparentes, éticas y responsables con esta potente herramienta y deben mantener una estrecha vigilancia para mitigar o compensar posibles efectos negativos[73].

El riesgo de la desinformación

La desinformación puede originarse en la IA generativa: por modelos mal entrenados que dan lugar a datos inexactos o sesgados[74], por el uso indebido de la misma y por ataques maliciosos

71 En parte del Considerando 56 LIA se expone que "El despliegue de sistemas de IA en el ámbito educativo es importante para fomentar una educación y formación digitales de alta calidad y para que todos los estudiantes y profesores puedan adquirir y compartir las capacidades y competencias digitales necesarias, incluidos la alfabetización mediática, y el pensamiento crítico, para participar activamente en la economía, la sociedad y los procesos democráticos." Se refieren también a la alfabetización los Considerandos 91, 165 y los arts.66 c) y 95.2 c) LIA.

72 FRANGANILLO, J. "La inteligencia artificial...op. cit. pág.14.

73 IBIDEM., pág.14.

74 GOLDMAN SACHS, "Principales...op.cit.

(deepfakes o fakes news) con información engañosa, lo que contribuiría a la difusión de noticias falsas y desinformación[75].

La toxicidad

Los modelos de IA generativa pueden propagar, al reflejar el lenguaje de la red, contenidos tóxicos (por ejemplo, blasfemias y contenido sexual explícito) [76].

El riesgo en relación a la seguridad

Para ANNEMANS [77]son posibles problemas y riesgos en relación a la seguridad de la IA generativa: el desbordamiento de datos, la fuga de la propiedad intelectual y la confidencialidad, el entrenamiento de datos, el almacenamiento de datos, el cumplimiento, los datos sintéticos, las fugas accidentales, el uso indebido de la IA y los ataques maliciosos. La IA generativa necesita una gran cantidad de datos (puede incorporar datos de cualquier tipo -incluidos con información confidencial y/o privada-), que se van incrementando para conseguir un mayor aprendizaje y mejora del modelo. En el entrenamiento de datos (entrenamiento de algoritmos), podrían desvelarse involuntariamente datos

75 El Reglamento (UE) 2024/1083 del Parlamento Europeo y del Consejo, de 11 de abril de 2024, por el que se establece un marco común para los servicios de medios de comunicación en el mercado interior y se modifica la Directiva 2010/13/UE (Reglamento Europeo sobre la Libertad de los Medios de Comunicación), DOUE, nº 1083, de 17 de abril de 2024 recoge la lucha contra la desinformación en los Considerandos 4,6,14,51,53 y 56 y el art.19.1.c). Por su parte en el Estudio, la política europea frente a los deepfakes de julio de 2021, se define la desinformación como "difusión consciente, normalmente encubierta, de información engañosa con el objetivo de perjudicar el debate público, los procesos democráticos, la economía abierta o la seguridad nacional" (pág.XIV) y reconoce modalidades de la misma (pág.23). MUÑOZ VELA, J.M, *Retos...* op.cit.págs.93 y ss., sobre el concepto de desinformación.

76 Documento de debate Generative AI:Implications for Trust and Governance, 4, publicado en junio del 2023, págs.10 y 11.

77 ANNEMANS,R.,"Seguridad de la IA generativa: 8 riesgos que debes conocer". GlobalSing by GMO, 4 de diciembre del 2023 https://www.globalsign.com/es/blog/8-riesgos-de-seguridad-de-la-inteligencia-artificial-generativa (recuperado el 10 de mayo del 2024).

confidenciales, afectando a privacidad. Datos, que pueden almacenarse por terceros, que en caso de ser confidenciales pueden usarse indebidamente o filtrase si no están suficientemente protegidos (por ejemplo con elementos de cifrado y controles de acceso). No se puede olvidar el necesario cumplimiento de las correspondientes normas de datos. Por otra parte, los datos sintéticos, a través de sus patrones o detalles podrían llevarnos a la identidad o características sensibles. Pueden darse fugas accidentales, por ejemplo, en los modelos con base en textos o imágenes que de forma involuntaria incluyan información de datos de entrenamiento que no deberían revelarse (información personal) o datos comerciales confidenciales. No podemos olvidar el gran valor de la información.

El riesgo para la privacidad y confidencialidad

Uno de los riesgos para la privacidad como reconoce el documento de debate Generative AI:Implications for Trust and Governance, 4, publicado en junio del 2023, es que los modelos "memoricen" al por mayor un registro de datos específico y cuando se les consulta lo repliquen, sobre todo cuando es información sensible (por ejemplo datos médicos). Esto también afecta a los derechos de autor y confidencialidad: por ejemplo, un empleado de Samsung que de manera involuntaria filtra información sensible cuando pega el código fuente, confidencial y protegido por derechos de autor, en ChatGPT comprobando si hay errores y optimizando el código[78].

Sobre la aplicación de la normativa de datos del Reglamento General de Protección de Datos, recientemente se han dado unas orientaciones del SEPD en el Informe, "La IA generativa y el EUDPR. Primeras orientaciones del SEPD para garantizar el cumplimiento de la protección de datos al utilizar sistemas de IA", de 3

78 Pág.10 del Documento.

de junio de 2024. Señalan estas orientaciones (que se entienden sin perjuicio de la LIA[79]), entre otras cosas, a que el uso masivo de datos disminuya: "El uso de grandes cantidades de datos para entrenar un sistema de IA generativa no implica necesariamente una mayor eficacia o mejores resultados. El diseño cuidadoso de conjuntos de datos bien estructurados, que se utilicen en sistemas que prioricen la calidad sobre la cantidad, siguiendo un proceso de entrenamiento debidamente supervisado y sometido a un seguimiento periódico, es esencial para lograr los resultados esperados, no solo en términos de minimización de datos, sino también cuando se trata de la calidad del resultado y la seguridad de los datos"[80]. También se reconoce la dificultad de erradicar los datos inexactos[81].

79 Son algunas de sus conclusiones que, "La supervisión periódica y la aplicación de controles en todas las etapas pueden ayudar a verificar que no hay tratamiento de datos personales, en los casos en que el modelo no está destinado a ello" (pág.8). "Es responsabilidad de la EUI gestionar adecuadamente los riesgos relacionados con el uso de sistemas de IA generativa. Los riesgos para la protección de datos deben identificarse y abordarse a lo largo de todo el ciclo de vida del sistema de IA generativa. Esto incluye una supervisión periódica y sistemática para determinar a medida que evoluciona el sistema, si los riesgos ya identificados están empeorando o si están apareciendo nuevos riesgos. La comprensión de los riesgos relacionados con el uso de IA generativa aún está en curso, por lo que es necesario mantener un enfoque vigilante con respecto a riesgos emergentes no identificados. Si se identifican riesgos que no pueden mitigarse por medios razonables, es el momento de consultar al SEPD" (págs.12 y ss.)." Las instituciones de la UE deben proporcionar a las personas físicas toda la información exigida en el Reglamento cuando utilicen sistemas de IA generativa que traten datos personales. La información facilitada a las personas deberá actualizarse cuando sea necesario para mantenerlas debidamente informadas y en control de sus propios datos" (pág.25)." Cuando se prevean sistemas de IA generativa para apoyar los procedimientos de toma de decisiones, las instituciones de la UE deberán considerar detenidamente la posibilidad de ponerlos en funcionamiento si su uso plantea dudas sobre su legalidad o su potencial de constituir decisiones injustas, poco éticas o discriminatorias" (pág.27)." La falta de información sobre los riesgos de seguridad ligados al uso de sistemas de IA generativa y su posible evolución obliga a las IUE a extremar la precaución y a realizar una planificación detallada de todos los aspectos relacionados con la seguridad informática, incluyendo la monitorización continua y el soporte técnico especializado. Las IUE deben ser conscientes de los riesgos derivados de los ataques de terceros malintencionados y de las herramientas disponibles para mitigarlos" (pág.35).

80 Pág.20, añade que "a medida que las tecnologías de IA avanzan rápidamente, las instituciones de la UE deben considerar cuidadosamente cuándo y cómo utilizar la IA generativa de manera responsable y beneficiosa para el bien público. Todas las etapas del ciclo de vida de una solución de IA generativa deben operar de conformidad con los marcos jurídicos aplicables, incluido el Reglamento, cuando el sistema implique el tratamiento de datos personales "(pág.7).

81 Así afirma el Informe del SEPD "La IA generativa y el EUDPR. Primeras orientaciones del SEPD para garantizar el cumplimiento de la protección de datos al utilizar sistemas de

El riesgo para la ciberseguridad

Para MUÑOZ VELA la seguridad lógica o informática tiene impacto en el mundo físico para personas, instalaciones (especialmente alto en las infraestructuras críticas y servicio esenciales), empresas y Gobiernos. Señala el autor el crecimiento exponencial de los ciberataques cuantitativa y cualitativamente, contra la reputación, la democracia, la influencia política y el ciberterrorismo[82]. Destacando en el ámbito de la IA generativa la posible aplicación de las técnicas de Deep Fakes para realizar el ciberataque del spear phishing o phishing selectivo y concretamente el "whaling" o "fraude del CEO"[83]. Ciberataques con co-

IA", de 3 de junio de 2024 que "A pesar de los esfuerzos por garantizar la exactitud de los datos, los sistemas generativos de IA siguen siendo propensos a resultados inexactos que pueden repercutir en los derechos y libertades fundamentales de las personas. Aunque los proveedores están implantando sistemas de información avanzados para garantizar que los modelos utilicen y generen datos precisos, las IUE deben evaluar cuidadosamente la precisión de los datos a lo largo de todo el ciclo de vida de los sistemas de IA generativa y plantearse el uso de dichos sistemas si no se puede mantener la precisión" (pág.22). Véase por otra parte, "El Comisionado de Hamburgo para la protección de Datos y la Libertad de Información pública en documento en el que se analiza los grandes Modelos de Lenguaje desde el punto de vista de la protección de datos, Boletín de julio del 2024", Lks, noticias, (https://www.lksnext.com/es/noticias_boletin/el-comisionado-de-hamburgo-para-la-proteccion-de-datos-y-la-libertad-de-informacion-publica-un-documento-en-el-que-analiza-los-grandes-modelos-de-lenguaje-desde-el-punto-de-vista-de-la-proteccion-de-d/) (rescuperado el 2 de agosto del 2024).

82 MUÑOZ VELA, J.M, *Retos*...op.cit.pág152 especifica, ataques distribuidos de denegación de servicios (DDos), mediante redes botnet (Como "Mirai atack"), ataques de ransomware, ataques incesantes de ciberespionaje mediante virus, especialmente para acceder y sustraer información confidencial , como "Moonlight Maza", "Titan Rain", "Duq"," Flame", "Red October" o "Gauss", entre muchos otros (ej. Cisco informó en 2018 que se bloquearon 7 billones de amenazas en nombre de sus clientes").

83 IBIDEM., págs.152 y ss., señala el autor que está este ciberataque dirigido a altos cargos para obtener información confidencial de una organización o dinero y, el "spear phishing basado en inteligencia artificial". Como expone el autor los delincuentes, que incluso forman parte de organizaciones criminales (Crimen as a Service), se benefician de estos sistemas para rastrear las redes (datos rastreables que se comparten por internet: dirección electrónica, imágenes y la voz) y encontrar información útil sobre la persona a suplantar, analizarla e imitar su lenguaje, su estilo de comunicación o su timbre de voz (por ejemplo dirigen una orden a un subordinado de una entidad, suplantando a la persona y falsificando las comunicaciones y su contenido). Por ejemplo, una empleada de una firma que tiene su sede en Hong Kong (en algunos medios hablan de empleado), transfirió 25 millones de dólares a unos estafadores tras recibir esa instrucción por su director financiero en una videollamada con otros profesionales. Realmente no era una llamada real sino una

rreos electrónicos de phishing y código malicioso que incluso poniendo filtros para impedirlos se los pueden saltar[84].

El uso de la IA como instrumento para la comisión de actos ilícitos o delictivos, seguirá creciendo, estudiando el comportamiento del sistema y humano ("la parte débil"). Es necesario por ello, reforzar la seguridad de la IA generativa y controlarla ya que puede ser instrumento para el ciberataque (dada la información que guarda)[85] y al mismo tiempo la IA se convierte en instrumento al servicio de la ciberseguridad[86]. Pudiendo desempeñar en materia de ciberseguridad un papel destacado[87].

réplica creada por los estafadores (LALCHAND, S. et al. (Centro de Servicios Financieros de Deloitte), "Se espera que la IA generativa aumente el riesgo de deepfakes y otros fraudes de la banca", Deloitte, Servicios Financieros, 29 de mayo del 2024, https://www2.deloitte.com/us/en/insights/industry/financial-services/financial-services-industry-predictions/2024/deepfake-banking-fraud-risk-on-the-rise.html)) (recuperado el 2 de agosto del 2024).

84 Documento de debate Generative AI:Implications for Trust and Governance, 4, publicado en junio del 2023, pág.16.

85 Es necesario actualizar la legislación (SÁNCHEZ, L., "Escrivá advierte que la estrategia de IA necesita de entornos seguros y anuncia una nueva ley integral de ciberseguridad", Economist & Jurist, 7 de junio del 2024 (recuperado en 20 de julio del 2024)).

86 MUÑOZ VELA, J.M, *Retos*...op.cit.pág.149, afirma que "La IA puede servir para garantizar o mejorar la seguridad, especialmente mediante medidas y controles preventivos , de efectivos y reactivos de gestión, contención y mitigación y para atentar con la misma con absoluta precisión, personalización y efectividad.".

87 "La inteligencia artificial generativa puede desempeñar un papel valioso en el campo de la ciberseguridad de varias maneras interesantes: 1. Simulación de amenazas: La IA generativa puede crear escenarios de ciberataques simulados que ayudan a entrenar a los sistemas de defensa cibernética y a los profesionales en la identificación y mitigación de riesgos. Esto permite a las organizaciones prepararse mejor para ataques reales y entender mejor las tácticas que los atacantes podrían utilizar. 2. Generación de Datos de Prueba:Puede generar grandes cantidades de datos de red realistas pero ficticios que pueden ser usados para entrenar modelos de detección de intrusos sin comprometer datos reales sensibles. Esto es especialmente útil para empresas que necesitan cumplir con regulaciones estrictas de privacidad.3.Análisis de Comportamiento Anómalo: La IA generativa puede ayudar a modelar lo que se considera "normal" en un red y luego generar comportamientos que se desvíen de esta norma, lo que es crucial para sistemas de detección de anomalías.4.Fortalecimiento de la Autenticación: En la autenticación, técnicas generativas pueden ser utilizadas para crear métodos de verificación biométrica más complejos y seguros, como la generación de voz o imagen que ayuden a mejorar los sistemas de reconocimiento facial o de voz.5.Mejora de los Sistemas de Respuesta e Incidentes: Al integrar IA generativa, los sistemas de respuesta a incidentes pueden simular diversas estrategias de respuesta a ataques, permitiendo a las organizaciones evaluar y optimizar sus tácticas y protocolos de respuesta antes de que un

Riesgos psicológicos

En el Considerando 5 de la LIA se alude expresamente a los riesgos psicológicos, afirma que "dependiendo de las circunstancias relativas a su aplicación, utilización y nivel de desarrollo tecnológico concretos, la IA puede generar riesgos y menoscabar los intereses públicos y los derechos fundamentales que protege el Derecho de la Unión. Dicho menoscabo puede ser tangible o intangible e incluye los perjuicios físicos, psíquicos, sociales o económicos". La IA generativa puede tener un impacto psicológico y emocional. Por ejemplo, el posible sufrimiento, malestar de familiares y allegados, al escuchar una voz clonada y ver la reproducción de la imagen de una persona fallecida o desaparecida[88]. Impacto que podría generar daños psicológicos tanto si se usa de forma normal (es algo impactante) como de manera imprudente o manipulada. Como afirma FRANGANILLO[89] toda manipulación tiene un impacto psicológico, por eso son peligrosos los deefakes y las alucinaciones o fallos. No sabemos el alcance de las consecuencias. Se ha llegado a producir un suicidio inducido por inteligencia artificial, un chatbot ("Eliza") basado

incidente real ocurra" RESPUESTA de Chat GPT a prompt: Puede la IA generativa servir de ciberseguridad, que le planteé el 10 de mayo del 2024 a las 15,26. Por su parte, Microsoft Security establece medidas de protección frente a las amenazas de ciberseguridad con IA generativa en su guía: IA generativa, la ventaja de los defensores, dirigida a los directivos de seguridad de la información (CISO) y a los profesionales en ciberseguridad que buscan colaborar con directivos de toda su organización y asegurarse de que la empresa aprovecha la IA generativa, LinkedIn, Microsoft (recuperado en 6 de mayo de 2024).

88 En el artículo "Las empresas de inteligencia artificial ofrecen ya el servicio de recrear a seres queridos fallecidos e interactuar con ellos",20 minutos,20 bits, 3 de diciembre del 2023,(https://www.20minutos.es/tecnologia/empresas-inteligencia-artificial-ofrecen-servicio-recrear-seres-queridos-fallecidos-interactuar-con-ellos-5195903/) (recuperado el 1 de junio del 2024) se muestra como se ofrece este servicio de recrear seres queridos fallecidos e interactuar con ellos, sus beneficios y problemas (entre ellos, que se quiera maximizar las ganancias por parte de la empresa con ofertas "adictivas" como cobrar cada vez que lo utilicen en vez de usar tarifa fija) y el debate ético que se plantea. Muestra como ya hay gente preparando su avatar posterior, preparan su "yo" virtual para después de la muerte. Véase MUÑOZ VELA, J.M, *Retos*...op.cit.pág.76.

89 FRANGANILLO, J. "La inteligencia artificial...op. cit. pág.14.

en tecnología GPT-J[90]. El Estudio, la política europea frente a los deepfakes, de julio del 2021 reconoce entre los daños psicológicos: la extorsión, la sextorsión, la difamación, la intimidación, el acoso escolar y socavar la confianza[91]. Otro ejemplo, sería como dice MUÑOZ VELA[92], el previsible fuerte impacto psicológico del desempleo temporal, y sobre todo, la exclusión permanentemente del mundo laboral, ante el cambio que está provocando y provocará la IA en el mercado laboral.

Riesgos por errores y alucinaciones

Los errores y alucinaciones[93] (cuando son vividos y adoptan una antropomorfización) son riesgos de la IA generativa. Chat GPT, comete errores, tiene mayores dificultades para tareas como la lógica, las matemáticas y el sentido común. Crea respuestas erróneas muy convincentes, "fiables", por ejemplo, a preguntas médicas. Ha creado falsas historias de acoso sexual y código de software susceptible de vulnerabilidades. No perdiendo de vista que cualquier vulnerabilidad de un modelo base corre el riesgo de heredarse en los modelos derivados de él.[94]

90 SOTO ARAMENDARIZ, S "Primer suicidio inducido por inteligencia artificial: algo que temer, 4 de abril de 2023 (https://observatorioblockchain.com/ia/primer-suicidio-inducido-por-inteligencia-artificial-algo-que-temer) (recuperado el 28 de mayo del 2024), aunque la IA también ayuda a detectar los posibles intentos de suicidio y evitarlos (PUFPAFF, M. ¿ La tecnología como fuerza para el bien? ¿Cómo se está utilizando la inteligencia artificial para prevenir los suicidios en China?, Razón y fe, Tomo 282, nº1447, 2020, págs.205 y ss.).

91 Págs.30., reconoce también el Informe daños financieros y sociales (págs.30 y ss).

92 MUÑOZ VELA, J.M, *Retos*...op.cit.pág.44.

93 En el documento de debate Generative AI:Implications for Trust and Governance, 4, publicado en junio del 2023 se considera más correcto el uso del término "confabulación" o pastiche (pág.9). Considera preocupantes estos problemas en modelos de cimentación por su diseño para un uso amplio y general.

94 Documento de debate Generative AI:Implications for Trust and Governance, 4, publicado en junio del 2023 , pág.9.

Riesgos en relación a las cuestiones éticas. Códigos de conducta y alienación de valores

La IA generativa puede mejorar la productividad y ofrecer beneficios, pero como se ha visto, implica riesgos éticos y sociales (¿Hasta qué punto es ético "revivir" a las personas fallecidas?, la manipulación del público generando confusión y desinformación y el contenido sesgado o engañoso, con los deepfakes, por ejemplo) que afectan a aspectos esenciales como el empleo, la democracia, la cultura, el sistema económico, la privacidad, la propiedad intelectual, la intimidad, la no discriminación (creada o acentuada por la IA), etc. Los contenidos son tan realistas que pueden confundir socialmente, por ello como establece FRANGANILLO[95], es exigible un código de conducta, una concienciación sobre un uso ético y responsable, para el bien común, en iniciativas con impacto social, como el proyecto OpenAI, el observatorio OdiseaIA o el programa AI for Social Good, de Google. Un uso transparente y responsable de la tecnología que respete los valores fundamentales de la sociedad.

La seguridad de la IA generativa se asocia a menudo, según el documento de debate Generative AI:Implications for Trust and Governance, 4, publicado en junio del 2023[96] con el concepto de alineación de valores (alineados con los valores y objetivos humanos para evitar que hagan daño a sus creadores humanos). Se formulan instrucciones para que la IA generativa alcance determinados "objetivos" que pueden estar mal especificados o representados. Explica el Informe que una función objetivo para los asistentes de IA debe ser que sea "útil" o "inofensivo", conceptos difíciles de definir y especificar y de determinar su compensación. Pone el ejemplo, de que la insistencia en evitar el "daño" puede llevar a respuestas "seguras" que quizás no sean

95 FRANGANILLO, J. "La inteligencia artificial...op. cit. pág.11

96 Documento de debate Generative AI:Implications for Trust and Governance, 4, publicado en junio del 2023, pág.19

valiosas para el usuario, pero, por otra parte, dar más importancia a ser útil puede hacer que el sistema ofrezca respuestas tóxicas que causen daños. Reconoce que se puede mitigar este problema, basándose en el Aprendizaje por Refuerzo a través de la Retroalimentación Humana (RLHF).

III.- RIESGOS DE LA IA GENERATIVA QUE HAY QUE TENER EN CUENTA EN RELACIÓN A LA LIA

Como expone JARQUES[97], la idea principal de la LIA es regular la IA en función de su capacidad de causar daño a la sociedad con un enfoque "basado en el riesgo", a mayor riesgo más estrictas son las reglas. La normativa establece obligaciones para la IA en función de los riesgos potenciales y su nivel de impacto, dividiéndose los sistemas de IA en: Riesgo inaceptable (prohibidos), alto (requiere estrictas medidas), limitado y mínimo [98].

97 JARQUES, A., "El futuro Reglamento Europeo de Inteligencia Artificial", *Actualidad Jurídica Aranzadi,* nº1003, 2023, Editorial Aranzadi, (BIB 2024/278) (rescatada en 27 de mayo del 2024). Afirma que los dos últimos tienen regulaciones y medidas menos estrictas, sus obligaciones de transparencia son más leves (señala la autora la obligación de divulgar que el contenido se generó mediante IA).

98 La idea podría seguir siendo válida tras la redacción definitiva del Reglamento, ya que continúa la IA prohibida y la de alto riesgo. Se puede considerar con menor riesgo ("limitado") a la IA de riesgos sistémicos (siempre que no esté incluída en la prohibida y la de alto riesgo), y también la IA presenta en determinadas ocasiones un riesgo mínimo fuera de los tres casos mencionados. En esta línea establece el Considerando 26 LIA que"con el fin de establecer un conjunto proporcionado y eficaz de normas vinculantes para los sistemas de IA, es preciso aplicar un enfoque basado en los riesgos claramente definido, que adapte el tipo y contenido de las normas a la intensidad y el alcance de los riesgos que puedan generar los sistemas de IA de que se trate. Por consiguiente, es necesario prohibir determinadas prácticas de IA que no son aceptables, definir los requisitos que deben cumplir los sistemas de IA de alto riesgo y las obligaciones aplicables a los operadores pertinentes, así como imponer obligaciones de transparencia a determinados sistemas de IA".

III.1.- LA IA GENERATIVA COMO IA PROHIBIDA

Un punto de partida en común, para toda clase de sistema de IA, es el establecimiento de prácticas prohibidas, aplicables a la IA de uso general, y por tanto a los grandes modelos de IA generativa. Se están de esta forma limitando a priori posibles riesgos inaceptables del uso de una IA generativa, evitándolos de forma taxativa.

En el artículo 5 LIA se recogen las prácticas de IA que quedan prohibidas y también las excepciones a las mismas. Se prohíbe la introducción en el mercado, la puesta en servicio o la utilización de un sistema de IA considerado prohibido por dicha norma.

Dentro de estas prohibiciones, se observa, que algunas pueden afectar directamente a la IA generativa y otras, entiendo, podrían hacerlo, solo si se usan para el entrenamiento de otras tecnologías prohibidas, o en apoyo a las mismas.

En primer lugar, podrían considerarse como prácticas prohibidas que afectan directamente a la IA generativa, principalmente:

1º- La "que se sirva de técnicas subliminales que trasciendan la conciencia de una persona o de técnicas deliberadamente manipuladoras o engañosas con el objetivo o el efecto de alterar de manera sustancial el comportamiento de una persona o un colectivo de personas, mermando de manera apreciable su capacidad para tomar una decisión informada y haciendo que tomen una decisión que de otro modo no habrían tomado, de un modo que provoque, o sea razonablemente probable que provoque, perjuicios considerables a esa persona, a otra persona o a un colectivo de personas" (art.5.1 a) LIA). En el artículo inicialmente, la necesidad de alteración sustancial del comportamiento parece exigible a ambos supuestos, pero podría interpretarse que el primer supuesto "que se sirva de técnicas subliminales que transciendan la conciencia de una persona" debe ser un supuesto de prohibición

en sí mismo (después de engañosas no lleva coma lo que podríamos interpretar que el resto de matiz del artículo se refiere al segundo supuesto: técnicas deliberadamente manipuladoras o engañosas) sin necesidad de alterar sustancialmente el comportamiento a la hora de tomar la decisión. PLAZA PENADÉS[99] ya consideró y comparto, que realmente se debe prohibir toda IA que se sirva de técnicas subliminales que transciendan la conciencia de una persona, independientemente de que sea determinante o no en la decisión, por "el principio de que "los derechos fundamentales están por encima de la tecnología" y "porque los sistemas IA que se sirvan de técnicas subliminales que trasciendan la conciencia de una persona suponen una intromisión ilegítima al honor (en sentido inmanente, la concepción que una persona tiene de sí misma) y a su más estricta y pura intimidad". El autor manifiesta que esa primera prohibición podía haber sido un "neuroderecho" y redactarse así:"Quedan prohibidos los sistemas IA que tengan como finalidad acceder y trascender la conciencia de una persona, aunque sea simplemente para acceder a dicha conciencia. Revestirá especial gravedad si además pretenden alterar o alteran la voluntad cognitiva de las personas, salvo que sea por razones médicas y se realice con la supervisión de un médico especialista responsable".

Es evidente que a través de las imágenes, audios, videos y textos sintéticos creados por la IA generativa, como se expuso, se puede de forma subliminal transcender la conciencia

99 PLAZA PENADÉS, J., Dossier, "Las claves de la futura Ley de Inteligencia Artificial Europea", Aranzadi La Ley, Navarra, mayo, 2023, pág.15. Para el autor debe prohibirse la IA que afecte a los derechos fundamentales básicos de las personas: "Por tanto, podemos colegir que la principal prohibición es que un sistema IA infrinja o no garantice la observancia o cumplimiento de los derechos fundamentales básicos de las personas, lo que formulo siempre bajo el principio "los derechos fundamentales prevalecen sobre la tecnología", que debemos de aplicar a todo el desarrollo de Internet en lo que se refiere al necesario respeto de los derechos de honor, intimidad, propia imagen, protección de datos.... Principio que obviamente también resulta extensible y aplicable a todos los sistemas de Inteligencia Artificial" (pág.14).

de una persona, o deliberada, manipular y engañar con el objetivo o el efecto de alterar de manera sustancial (es determinante en su actuación) el comportamiento de una persona o colectivo de personas que impide tomar una decisión informada y le lleva a tomar otra que provoca o puede provocar perjuicios considerables a la persona o colectivo de personas que la adoptan. En el momento que das una información, imagen o vídeo falsos o manipulados (deepfakes, fake news,deep voice), cuando no se puede distinguir fácilmente la realidad de la ficción, el uso en este sentido está prohibido.

2º- La "que explote alguna de las vulnerabilidades de una persona física o un determinado colectivo de personas derivadas de su edad o discapacidad, o de una situación social o económica específica, con la finalidad o el efecto de alterar de manera sustancial el comportamiento de dicha persona o de una persona que pertenezca a dicho colectivo de un modo que provoque, o sea razonablemente probable que provoque, perjuicios considerables a esa persona o a otra" (art.5.1 b) LIA).

Los contenidos creados por la IA generativa podrían utilizarse para explotar vulnerabilidades de una persona o colectivo específico de personas (por edad, discapacidad, situación social o económica específica), a través de textos, vídeos o imágenes falseadas dirigidas a ella o a un colectivo, haciéndole creer lo que no es, pero estas personas no son capaces de discernir, creando confusión, desinformación y manipulación. Todo ello, con el objetivo o efecto de alterar sustancialmente (es determinante en su actuación) su comportamiento esa persona u otra que pertenece al colectivo, provocando o siendo razonable que provoque perjuicios considerables a las mismas.

En estas prohibiciones como se ha visto, la ética está en la esencia de las mismas: la protección de las personas vulnerables y la

no manipulación del comportamiento de las personas que deben tomar decisiones libre y conscientemente[100].

En relación al resto de supuestos de prácticas prohibidas (con el fin de evaluar o clasificar a personas físicas o a grupos de personas, para realizar evaluaciones de riesgo de personas, base de reconocimiento facial, inferir emociones e IA de categorización biométrica), normalmente, el Sistema de IA principal no es una IA generativa. Por ejemplo, en materia de reconocimiento facial y de inferir emociones, habitualmente la base son otros sistemas, pero se puede utilizar la IA generativa para el entrenamiento de estos generando imágenes con rostros, o con rostros que tengan diferentes emociones. Podrían así ser útiles para crear datos sintéticos con los que entrenar estos sistemas de reconocimiento facial y de emociones, especialmente en situaciones donde los datos reales son limitados o cuando se necesita preservar la privacidad. En resumen, mientras que la IA generativa puede apoyar indirectamente las tareas de reconocimiento de emociones a través de la generación de datos para el entrenamiento, el trabajo directo de analizar y clasificar emociones se realiza mediante técnicas de IA no generativas especializadas en reconocimiento de patrones y aprendizaje automático. Entiendo que, si estas están prohibidas, también lo está el uso de una inteligencia generativa a su servicio.

La Comisión establecerá directrices sobre la aplicación del Reglamento sobre las prácticas prohibidas a que se refiere el art. 5 (art.96.1 b) LIA). Lo establecido en el art.5 LIA no afectará a las prohibiciones aplicables cuando una práctica de IA infrinja otras disposiciones de Derecho de la Unión (art.5.8 LIA).

La posible inclusión directa de los deepfakes como IA prohibida o, como se verá IA de alto riesgo, aparecía ya en el Estudio, la política europea frente a los deepfakes, de julio del 2021 que planteaba: "Prohibir determinadas aplicaciones: Considerando el

100 En el Considerando 29 de la LIA se recogen expresamente estas técnicas de manipulación.

potencial impacto negativo de aplicaciones específicas de deepfakes (p. ej., la pornografía deepfake no consentida), las obligaciones de transparencia por sí solas parecen insuficientes para hacer frente a esos efectos negativos. Por lo tanto, la Comisión proponía como opción prohibir determinados tipos de aplicaciones y usos concretos de esta tecnología"[101].

III.2.- LA IA GENERATIVA COMO IA GENERAL CON RIESGOS SISTÉMICOS

La IA generativa se incluye en los modelos de IA de uso general, modelos que en determinados casos pueden clasificarse como modelo de IA de uso general con riesgo sistémico. Aparece así uno de los riesgos reconocidos en la LIA el riesgo sistémico, definido en su artículo 3 número 65, como "un riesgo específico de las capacidades de gran impacto de los modelos de IA de uso general (aquellas que igualan o superan las capacidades mostradas por los modelos de IA de uso general más avanzados-definición 64 art.3 LIA-), que tienen unas repercusiones considerables en el mercado de la Unión debido a su alcance o a los efectos negativos reales o razonablemente previsibles en la salud pública, la seguridad, la seguridad pública, los derechos fundamentales o la sociedad en su conjunto, que puede propagarse a gran escala a lo largo de toda la cadena de valor".

Dicho riesgo se da si el modelo de IA de uso general reúne alguna de las siguientes condiciones (art.51 .1 LIA) :a) tiene capacidades de gran impacto evaluadas a partir de herramientas y metodologías técnicas adecuadas, como indicadores y parámetros de referencia; b) con arreglo a una decisión de la Comisión, adoptada de oficio a raíz de una alerta cualificada del grupo de expertos científicos, tiene capacidades o un impacto equivalente a los establecidos en la letra a), teniendo en cuenta los criterios establecidos en el anexo XIII. En dicho Anexo XIII aparecen

101 Pág..61. ÁLVAREZ, P. y EGUILUZ, J. "El Reglamento de IA...op.cit.

criterios de carácter técnico, objetivo (para la clasificación de los modelos de IA de uso general con riesgo sistémico a que se refiere el art. 51[102]).

Además de la presencia de posibles capacidades de gran impacto, es indudable que la IA generativa puede tener efectos negativos reales o razonablemente previsibles en la salud pública, la seguridad, la seguridad pública, los derechos fundamentales o la sociedad en su conjunto[103]. Esto ha quedado

102 Señala el art.51.2 y 3 LIA respectivamente que "se presumirá que un modelo de IA de uso general tiene capacidades de gran impacto con arreglo al apartado 1, letra a), cuando la cantidad acumulada de cálculo utilizada para su entrenamiento, medida en operaciones de coma flotante, sea superior a 10 25" y que "La Comisión adoptará actos delegados de conformidad con el artículo 97 para modificar los umbrales a que se refieren los apartados 1 y 2 del presente artículo, así como para complementar los parámetros de referencia e indicadores en función de los avances tecnológicos, como las mejoras algorítmicas o la mayor eficiencia del hardware, cuando sea necesario, para que los umbrales reflejen el estado actual de la técnica". Añade su art.52.6 LIA que "La Comisión velará por que se publique una lista de modelos de IA de uso general con riesgo sistémico, que mantendrá actualizada, sin perjuicio de la necesidad de respetar y proteger los derechos de propiedad intelectual e industrial y la información empresarial confidencial o los secretos comerciales de conformidad con el Derecho de la Unión y nacional".

103 Es esencial en este punto el Considerando 110 de LIA que recoge de forma detallada dichos riesgos "Los modelos de IA de uso general pueden plantear riesgos sistémicos, por ejemplo, cualquier efecto negativo real o razonablemente previsible en relación con accidentes graves, perturbaciones de sectores críticos y consecuencias graves para la salud y la seguridad públicas, cualquier efecto negativo real o razonablemente previsible sobre los procesos democráticos y la seguridad pública y económica o la difusión de contenidos ilícitos, falsos o discriminatorios. Debe entenderse que los riesgos sistémicos aumentan con las capacidades y el alcance de los modelos, pueden surgir durante todo el ciclo de vida del modelo y se ven influidos por las condiciones de uso indebido, la fiabilidad del modelo, la equidad y la seguridad del modelo, el nivel de autonomía del modelo, su acceso a herramientas, modalidades novedosas o combinadas, las estrategias de divulgación y distribución, la posibilidad de eliminar las salvaguardias y otros factores. En particular, los enfoques internacionales han establecido hasta la fecha la necesidad de prestar atención a los riesgos derivados de posibles usos indebidos intencionados o de problemas en materia de control relacionados con la armonización con la intención humana no deseados, a los riesgos químicos, biológicos, radiológicos y nucleares, como las maneras en que las barreras a la entrada pueden reducirse, también para el desarrollo, el diseño, la adquisición o el uso de armas, a las cibercapacidades ofensivas, como las maneras en que pueden propiciarse el descubrimiento, la explotación o el uso operativo de vulnerabilidades, a los efectos de la interacción y el uso de herramientas, incluida, por ejemplo, la capacidad de controlar sistemas físicos e interferir en el funcionamiento de infraestructuras críticas, a los riesgos derivados del hecho que los modelos hagan copias de sí mismos o se «autorrepliquen» o entrenen a otros modelos, a las maneras en que los modelos pueden dar lugar a sesgos dañinos y discriminación que entrañan riesgos para las personas, las comunidades o las sociedades, a la

expuesto anteriormente al analizar ampliamente sus riesgos los cuales afectan o pueden afectar a todas las esferas antes mencionadas. Estos riesgos pueden propagarse a gran escala a lo largo de toda la cadena de valor. También queda reflejada esa posibilidad en el Considerando 136 de la LIA al hablar de las obligaciones impuestas a los proveedores y a los responsables del despliegue de determinados sistemas de IA (detectar, divulgar que el resultado es artificial y etiquetarlo), se refiere expresamente a los riesgos sistémicos que pueden surgir de la divulgación de contenidos generados o manipulados de manera artificial.

III.3.- LA IA GENERATIVA COMO IA DE ALTO RIESGO

Una IA generativa puede ser o convertirse en una IA de alto riesgo sujeta a la más estricta regulación a la que está sometida dicha IA en la LIA.

Los sistemas de IA de alto riesgo[104] (art.6.1 a) y b) LIA) son los que están destinados a utilizarse como componentes de seguridad de un producto (que entre en los ámbitos de aplicación de los actos legislativos de armonización de la Unión de su anexo I) o son uno de esos productos sujetos ambos a una evaluación de la conformidad de terceros para su introducción en el mercado

facilitación de la desinformación o el menoscabo de la intimidad, que suponen una amenaza para los valores democráticos y los derechos humanos, al riesgo de que un acontecimiento concreto dé lugar a una reacción en cadena con efectos negativos considerables que podrían afectar incluso a una ciudad entera, un ámbito de actividad entero o una comunidad entera". Posteriormente, en el Considerando 111 de la LIA se muestra el establecimiento de una metodología para la clasificación de los modelos de IA de uso general como modelos de IA de uso general con riesgos sistémicos y en el 112, el procedimiento para la clasificación de un modelo de IA de uso general con riesgos sistémicos.

104 En La LIA estos sistemas se regulan en el capítulo III: en la sección 1 (arts.6.-7) se dispone su clasificación. En la sección 2 se recogen los requisitos de los Sistemas de las IA de alto riesgo (arts.8-15 LIA). En la sección 3 se recogen las obligaciones de los proveedores y responsables del despliegue de sistemas de IA de alto riesgo y otras partes (arts.16-27 LIA). En la sección 4, recoge las autoridades notificantes y los organismos notificados (arts.28-39) En la sección 5, Normas, evaluación de la conformidad, certificados, registros (arts.40-49 LIA).

o puesta en servicio, con arreglo a los mencionados actos legislativos de armonización de la Unión Europea, enumerados en el anexo I. También son sistemas de IA de alto riesgo, los sistemas que formen parte de cualquiera de los ámbitos recogidos en el Anexo III (art.6.2 LIA) pero deben plantear un riesgo importante de causar un perjuicio a la salud, la seguridad o los derechos fundamentales de las personas físicas, e influir sustancialmente en el resultado de la toma de las decisiones (a sensu contrario del art.6. 3 LIA que no considera a una IA de alto riesgo cuando no plantee un riesgo importante de causar un perjuicio a la salud, la seguridad o los derechos fundamentales de las personas físicas, también al no influir sustancialmente en el resultado de la toma de decisiones y cumpla cualquiera de las condiciones que establece el artículo, aunque siempre se considerarán de alto riesgo cuando el sistema de la IA elabore perfiles de personas físicas).

Las materias recogidas en el Anexo III son: biometría, en la medida en que su uso esté permitido por el Derecho de la Unión o nacional aplicable (sistemas de identificación y categorización biométrica y reconocimiento de emociones en los términos establecidos); infraestructuras críticas: sistemas de IA destinados a ser utilizados como componentes de seguridad ; educación y formación profesional; empleo, gestión de los trabajadores y acceso al autoempleo; acceso a servicios privados esenciales y a servicios y prestaciones públicos esenciales y disfrute de estos servicios y prestaciones; garantía del cumplimiento del Derecho, en la medida en que su uso esté permitido por el Derecho de la Unión o nacional aplicable; migración, asilo y gestión del control fronterizo, en la medida en que su uso esté permitido por el Derecho de la Unión o nacional aplicable y administración de justicia y procesos democráticos[105].

105 Sintetiza PLAZA PENADÉS, J., Dossier, "Las claves de...op. cit. págs.35 y 36: "Las que se utilicen para: -La identificación biométrica de personas y extraer conclusiones sobre sus características personales. Sin incluir los de verificación biométrica, como la autenticación. No son de alto riesgo los utilizados en ciberseguridad y para la protección

Por otra parte, el propio texto de la LIA nos detalla esos derechos fundamentales que no pueden sufrir un riesgo importante. Entre los derechos fundamentales protegidos por la Carta de los Derechos Fundamentales de la Unión Europea, a los efectos de clasificar un sistema como de alto riesgo, el Considerando 48 LIA incluye: el derecho a la dignidad humana, el respeto de la vida privada y familiar, la protección de datos de carácter personal, la libertad de expresión y de información, la libertad de reunión y de asociación, el derecho a la no discriminación, el derecho a la educación, la protección de los consumidores, los derechos de los trabajadores, los derechos de las personas discapacitadas, la igualdad entre hombres y mujeres, los derechos de propiedad intelectual, el derecho a la tutela judicial efectiva y a un juez imparcial, los derechos de la defensa y la presunción de inocencia, el derecho a una buena administración y los derechos específicos

---

de datos personales.-La educación o la formación profesional: determinan el acceso o admisión, distribución o evaluación de las personas a partir de pruebas realizadas.- El empleo, la gestión de los trabajadores y el acceso al autoempleo, en decisiones relativas a la iniciación, la promoción y la rescisión de contratos y la asignación personalizada de tareas y evaluación de personas en relaciones laborales.- Evaluar la calificación crediticia o solvencia de personas físicas, ya que deciden sobre el acceso a recursos financieros o servicios esenciales (vivienda, electricidad y servicios de telecomunicaciones). No son de alto riesgo los previstos para detectar fraudes en la oferta de servicios financieros.-Decidir si las autoridades deben denegar, reducir, revocar o reclamar ayudas y servicios.- Tomar decisiones o influir en la elegibilidad de las personas físicas para acogerse al seguro de enfermedad y de vida.- Evaluar y clasificar llamadas de emergencia de personas físicas o el envío o establecimiento de prioridades en el envío de servicios de primera intervención- La toma de decisiones de las autoridades policiales y judiciales en la prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de las sanciones.- Evaluar la fiabilidad de las pruebas en un proceso penal, elaborar perfiles, investigación o enjuiciamiento de infracciones penales o de ejecución de las sanciones.-Gestionar la migración, el asilo y el control fronterizo, ayudar a examinar y evaluar la veracidad de las pruebas, el seguimiento, la vigilancia o el tratamiento de datos personales en gestión de fronteras.- La administración de justicia y los procesos democráticos para ayudar a las autoridades judiciales u órganos administrativos a investigar e interpretar los hechos y el Derecho y a aplicar la ley. La utilización de estas herramientas no debe sustituir la toma de decisiones de los jueces ni su independencia, puesto que debe seguir siendo una actividad y una decisión humana; salvo, actividades administrativas accesorias (anonimización o seudonimización documentos o datos; comunicación entre personal; tareas administrativas, o la asignación de recursos).-Influir en el resultado de una elección o un referendo o en el comportamiento electoral en ejercicio del derecho a voto, a excepción de los sistemas para organizar, optimizar y estructurar campañas políticas desde el punto de vista administrativo y logístico".

de los menores. Añade que al evaluar la gravedad del perjuicio que puede ocasionar un sistema de IA, también en la salud y la seguridad de las personas, se debe tener en cuenta, además, el derecho fundamental a un nivel elevado de protección del medio ambiente[106].

La Comisión, previa consulta al Consejo Europeo de Inteligencia Artificial, dará directrices de aplicación práctica, en consonancia con el art.96 LIA y una lista exhaustiva de ejemplos prácticos de casos de uso de sistemas de IA que sean de alto riesgo y que no sean de alto riesgo (arts 6.5 LIA[107]). Hay que estar pues a la espera de dichas directrices y de la lista exhaustiva que dará luz a esta problemática. La lista de posibles IA de alto riesgo está abierta a futuras incorporaciones. La Comisión también puede adoptar actos delegados al objeto de aumentar o modificar la lista del Anexo III (art.97 en relación con el art.7LIA).

En el Anexo III LIA en su punto 8, en materia de administración de justicia y procesos democráticos, se recoge expresamente entre los sistemas considerados de alto riesgo: "b) Sistemas de IA destinados a ser utilizados para influir en el resultado de una elección o referéndum o en el comportamiento electoral de personas físicas que ejerzan su derecho de voto en elecciones o referendos. Quedan excluidos los sistemas de IA a cuyos resultados de salida no estén directamente expuestas las personas físicas, como las herramientas utilizadas para organizar, optimizar o estructurar campañas políticas desde un punto de vista

106 PUERTO MENDOZA, A., Derecho *digital*. *Fundamentos básicos*, Ediciones CEF.,2019 (págs.179 y ss.,200 y ss.) recoge junto al derecho al honor, la intimidad familiar y la propia imagen (art.18 CE y 12 de la Declaración universal de los Derechos Humanos (1948)) , el derecho al nombre (se menciona el anonimato y el problema de la suplantación y usurpación de la personalidad), el derecho a la libertad de expresión y de información (art.20 CE) y otros derechos fundamentales e instrumentales de los anteriores: el derecho a la inviolabilidad del domicilio, al secreto de las comunicaciones y a la autodeterminación informativa o derecho de protección de datos personales. Muestra un derecho constitucional de nueva generación: el derecho a la protección del propio entorno digital (págs.205 y ss.). Sobre los derechos digitales en general, págs.207 y ss.

107 A más tardar, como expresa el artículo el 2 de febrero del 2026.

administrativo o logístico"[108]. La IA generativa sería capaz de integrarse fácilmente en este supuesto, ya que podría a través de imágenes, voces o vídeos sintéticos ser utilizada para influir en el resultado de una elección o referéndum o en el comportamiento electoral de personas físicas que ejerzan su derecho de voto en elecciones o referendos. Por ejemplo, ha sido clara su finalidad de manipulación e influencia en las elecciones de Nigeria, Eslovenia, EEUU y México confundiendo al electorado[109].

108 En la STC núm. 76/2019 de 22 mayo (RTC 2019\76) se declaró inconstitucional y nulo el apartado 1 del art. 58 bis de la Ley Orgánica 5/1985, de 19 de junio, del régimen electoral general que permitía a los partidos políticos recopilar datos relativos a las opiniones políticas ya que como mostraba el motivo e) del recurso, las modernas técnicas de análisis de la conducta sobre la base del tratamiento masivo de datos y la inteligencia artificial permiten procedimientos complejos orientados a modificar, forzar o desviar la voluntad de los electores y sin que estos sean conscientes de ello.

109 En México, las redes sociales y las apps de mensajería instantánea (WhatsApp y Telegram) son los principales escenarios para la difusión de falsa información creando confusión entre la población. Por ejemplo, en un deep fake aparecía Claudia Sheinbaum, la candidata de la alianza "Seguimos haciendo historia", invitando a los electores a escribir en la boleta el nombre del presidente Andrés Manuel López Obrador, para que, supuestamente, continuara en la Presidencia. Por otra parte, Jorge Álvarez Máynez apareció en un vídeo en el que supuestamente estaba pidiendo a sus electores declinar su voto en favor de la candidata de la alianza Fuerza y corazón por México, Xóchitl Gálvez (CALDERÓN C.,"Elecciones México 2024: Deep fakes y fake news ganan en las votaciones", El financiero, 4 de junio del 2024,https://www.elfinanciero.com.mx/elecciones-mexico-2024/2024/06/04/deep-fakes-y-fake-news-marcaron-el-escenario-electoral/ (recuperado el 3 de agosto del 2024)). En EEUU son numerosos los ejemplos: algunos demócratas de New Hampshire recibieron llamadas automáticas que se habían generado por todos los Estados Unidos con la voz de Joe Biden instándoles a que se quedaran en casa en vez de ir a votar en las elecciones primarias del Estado(LINDSAY, J. M., "Elecciones 2024: La amenaza falsa a las elecciones del 2024", Council on Foreign Relations, 2 de febrero, 2024, https://www.cfr.org/blog/election-2024-deepfake-threat-2024-election (recuperado el 3 de agosto del 2024)). Utilizando imágenes de la campaña de Kamala Harris, se manipuló la voz en un vídeo haciendo declaraciones controvertidas que nunca hizo, aunque posteriormente el vídeo se etiquetó como una parodia, el daño ya estaba hecho (BANDARA, P., "Elon Musk compartió un vídeo engañoso de inteligencia artificial de la vicepresidenta Kamala Harris en X", PetaPixel, 29 de julio del 2024 (https://petapixel-com.translate.goog/2024/07/29/elon-musk-shared-misleading-ai-video-of-kamala-harris-on-x-twitter-deepfake-election-presidential-usa/?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=sc) (recuperado el 3 de agosto del 2024). Por otra parte, aparecía un vídeo de Taylor Swift manipulado, en una cuenta pro-Trump X que tiene más de un millón de seguidores (TENBARGE, K.,"Taylor Swift, deepfakes en X describe falsamente que apoya a Trump", NBC NEWS, 8 de febrero del 2024, https://www.nbcnews.com/tech/internet/taylor-swift-deepfake-x-falsely-depict-supporting-trump-grammys-flag-rcna137620 (recuperado el 3 de agosto del 2024)). Las elecciones en Nigeria de febrero de 2023 han sido objeto de numerosas desinformaciones, se difundió un clip de audio manipulado que implicaba falsamente a un

Debería considerarse, en esta situación, IA de alto riesgo y estar sujeta a los controles y obligaciones de transparencia de este tipo de IA.

Un sector doctrinal considera que la IA generativa y concretamente el deepfake debería haberse calificado directamente de alto riesgo[110]. Es más que evidente, como se expuso anteriormente, que pone en riesgo directo a numerosos de los derechos fundamentales reconocidos en el Texto de la LIA, y genera confusión y desinformación que pueden afectar a la capacidad de decisión en una elección o referéndum.

candidato presidencial en planes para manipular los resultados electorales. Así lo muestra ABDULKABBER, T.,"Funcionarios y famosos: así circuló la información falsa en las elecciones nigerianas", ijnet (red internacional de periodistas), Lucha contra la desinformación, 13 de abril del 2023 ,https://ijnet.org/es/story/funcionarios-y-famosos-as%C3%AD-circul%C3%B3-la-informaci%C3%B3n-falsa-en-las-elecciones-nigerianas) (recuperado el 3 de agosto del 2024), que afirma que se han registrado una serie de noticias falsas diseñadas para influir en la decisión de los votantes de todo el país. En las elecciones en Eslovenia se desacreditó al candidato, al difundirse unas grabaciones de audio falsificadas (unos días antes de las elecciones celebradas el 30 de septiembre del 2023). En una grabación, difundida en Facebook, se escuchaban dos voces: supuestamente, la de Michal Šimečka, líder del partido liberal Eslovaquia Progresista, y la de Monika Tódová, del diario *Denník N*. discutiendo la forma de amañar el proceso electoral, en parte comprando votos de la minoría romaní o gitana, marginada del país (y también se planteaba la subida del precio de la cerveza). En la otra grabación de audio de Šimečka, notificada por Meta al equipo de Demagog, difundida en Instagram, Šimečka aparece proponiendo duplicar el precio de la cerveza si vencía, siendo falsa. Estas grabaciones afectaron los resultados electorales y son un claro ejemplo del impacto de los deepfakes en la política (MEAKER, M., "Deepfakes en elecciones de Eslovaquia rearfiman que IA es un peligro para la democracia", Wired, negocios, 3 de octubre del 2023 https://es.wired.com/articulos/deepfakes-en-elecciones-de-eslovaquia-reafirman-que-ia-es-peligro-para-democracia) (recuperado el día 2 de agosto del 2024)).

110 GAMERO CASADO, E. "El enfoque europeo de la Inteligencia Artificial", *Revista de Derecho Administrativo*, nº 20, 2021, pág.279, considera que ya que no han sido prohibidos deberían estar sujetos a los sistemas de Inteligencia Artificial de Alto riesgo así afirma el autor al referirse a la lista de Sistemas de Inteligencia Artificial prohibidos "debe repararse en que esta lista, en realidad, es corta. Y en particular, que no rechaza la implantación de sistemas de la IA cuyos resultados resultan todavía incomprensibles para la mente humana, como el deep learning, las redes neuronales o los algoritmos de caja negra. Tales sistemas podrán utilizarse, sometiéndose, en su caso, a las reglas que analizamos a continuación ", esto último se refiere a las de los sistemas de IA de alto riesgo. Seguido por CASTILLO RAMOS-BOSSINI, S.E., "Regulación europea de la inteligencia artificial", *Nuevas fórmulas de prestación de servicios en la era digital*, dirección Juan Francisco Pérez Gálvez, Dyinson, 2023, pág.326.

En esta línea, estaba el Estudio, la política europea frente a los deepfakes, de julio del 2021, al exponer que[111] "teniendo en cuenta la larga lista de riesgos e impactos adversos asociados a los deepfakes, la Comisión arguye su clara afectación a los derechos fundamentales, la salud y a la seguridad, y hace un llamamiento para que las tecnologías de IA que creen deepfakes se categoricen como de alto riesgo. Categorizar estos sistemas como de alto riesgo, implicaría la aplicación de mayores obligaciones legales al proveedor de la tecnología, incluyendo la realización de evaluaciones de riesgos, la provisión de documentación, la supervisión humana y la garantía de la calidad de los conjuntos de datos de entrenamiento". Se ha mostrado a lo largo de este texto como la IA generativa puede afectar directamente y de forma grave a los derechos fundamentales, más allá de las decisiones electorales, pudiendo ser considerada de alto riesgo en otros supuestos y concretamente en materia de deepfake, en todo caso (si no se incluye en los casos de IA prohibida como se expuso). En dicho Estudio se recogen los daños psicológicos, financieros y sociales que generan los deepfakes incluyendo en esto últimos: la manipulación de los medios de comunicación, los daños a la estabilidad económica, al sistema judicial, al sistema científico, a la democracia, a las relaciones internacionales, la erosión de la confianza, la manipulación de las elecciones y los daños a la seguridad nacional[112]. Como confirma MUÑOZ VELA los deep fakes los han concebido distintos gobiernos como fuente de problemas de seguridad nacional en los últimos años[113].

Por otra parte, en los sistemas de IA relativos a ámbitos predefinidos especificados en el Reglamento, pueden darse casos específicos, que no entrañen un riesgo considerable de causar un perjuicio a los intereses jurídicos que se encuentran amparados en dichos ámbitos, al no influir sustancialmente en el resultado

111 ÁLVAREZ, P. y EGUILUZ, J. "El Reglamento de IA…op.cit., lo resumen.

112 Pág.IV.

113 MUÑOZ VELA, J.M, *Retos*…op.cit.pág.96.

de la toma de decisiones se entiende que el sistema de IA no afecta al fondo, ni por consiguiente al resultado de la toma de decisiones, ya sea humana o automatizada[114].

IV.- ESPECIAL REFERENCIA A LAS OBLIGACIONES DE CIBERSEGURIDAD Y TRANSPARENCIA Y LA ÉTICA

En conexión con las obligaciones de la IA generativa en la LIA, voy a destacar algunos puntos.

Ciberseguridad

En la Inteligencia Artificial de uso general de riesgo sistémico, en la que encuadramos a la IA generativa como se ha expuesto, la ciberseguridad hace acto de presencia. El art.55 LIA (sobre las obligaciones de los proveedores de modelos de IA de uso general con riesgo sistémico), establece en su punto 1 d) que entre las obligaciones de proveedores de modelos de IA de uso general con riesgo sistémico (además de las mencionadas en los arts. 53 y 54 LIA) está la de velar porque se establezca un nivel adecuado de protección de la ciberseguridad para el modelo de IA de uso general con riesgo sistémico y la infraestructura física del modelo[115].

114 El art.6.3 LIA muestra que un sistema así podría incluir situaciones en las que se cumplen cualquiera de las siguientes condiciones:" a) que el sistema de IA esté destinado a realizar una tarea de procedimiento limitada; b) que la tarea realizada por el sistema de IA esté destinada a mejorar el resultado de una actividad humana previa realizada; c) que el sistema de IA esté destinado a detectar patrones de toma de decisiones o desviaciones con respecto a patrones de toma de decisiones anteriores y no esté destinado a sustituir la valoración humana previamente realizada sin una revisión humana adecuada, ni influir en ella o d) que el sistema de IA esté destinado a realizar una tarea para una evaluación que sea pertinente a efectos de lo casos de uso enumerados en el anexo III.". Véase también el Considerando 53 LIA.

115 En el Considerando 115 se refleja esta idea.

También se alude a la ciberseguridad en la IA de propósito general con riesgo sistémico en el Considerando 114 al exigir a los proveedores de esta IA garantizar un nivel adecuado de protección en materia de ciberseguridad, independientemente de si los modelos se ofrecen como independientes o integrados en Sistemas de IA o en productos. En su Considerando 115, después de enumerar los posibles riesgos cibernéticos asociados al uso malintencionado o ataques[116], ofrece algunas medidas para fortalecer la ciberseguridad, reconociendo que debe tenerse en cuenta que "esa protección podría facilitarse asegurando los pesos, los algoritmos, los servidores y los conjuntos de datos del modelo, por ejemplo, mediante medidas de seguridad operativa para la seguridad de la información, medidas específicas en materia de ciberseguridad, soluciones técnicas adecuadas y establecidas y controles de acceso cibernéticos y físicos, en función de las circunstancias pertinentes y los riesgos existentes".

Se produce así un avance ya que en la Propuesta de Reglamento de 21 de abril de 2021 de la LIA la exigencia de ciberseguridad se centraba en los Sistemas de Inteligencia Artificial de Alto Riesgo[117]. Se recoge la exigencia de ciberseguridad en la LIA para los sistemas de alto riesgo principalmente en su art.15 (que regula requisitos de precisión, solidez y ciberseguridad). El art.15 exige la resistencia de los sistemas de IA de alto riesgo, en su apartado 5, ante intentos de alteración por parte de terceros no autorizados. El art.15.1 establece que "los sistemas de IA de alto riesgo se diseñarán y desarrollarán de modo que alcancen un nivel adecuado de precisión, solidez y ciberseguridad y funcionen de manera uniforme en esos sentidos durante todo su ciclo de vida". En su párrafo final el art.15 dispone que las soluciones técnicas encaminadas a garantizar la ciberseguridad de los sistemas de

116 Asociados "al uso malintencionado o a los ataques debe tener debidamente en cuenta las fugas accidentales de modelos, las divulgaciones no autorizadas, la elusión de las medidas de seguridad y la defensa contra los ciberataques, el acceso no autorizado o el robo de modelos".

117 MUÑOZ VELA, J.M, *Retos*...op.cit.pág.164

IA de alto riesgo serán adecuadas a las circunstancias y los riesgos pertinentes.

Entre las soluciones técnicas destinadas a subsanar vulnerabilidades específicas de la IA figurarán, como refleja el art.15.5 LIA, "según corresponda, medidas para prevenir, detectar, combatir, resolver y controlar los ataques que traten de manipular el conjunto de datos de entrenamiento («envenenamiento de datos»), o los componentes entrenados previamente utilizados en el entrenamiento («envenenamiento de modelos»), la información de entrada diseñada para hacer que el modelo de IA cometa un error(« ejemplos adversarios» o «evasión de modelos»), los ataques a la confidencialidad o los defectos en el modelo "[118]. Hay que recordar que el Estudio, la política europea frente a los deepfakes, de julio del 2021 mostraba su preocupación por la detección de deepfakes.

Pese a que en el texto final de la LIA se amplía la exigencia de la obligación de garantizar la ciberseguridad más allá de la IA de alto riesgo, avanzando así en relación a la Propuesta del 2021 (no hay que olvidar que la IA generativa puede ser IA de alto riesgo), sigue siendo insuficiente ya que la obligación de la ciberseguridad debería incluirse en una disposición común (al igual que las exigencias éticas) y ser exigible con unos mínimos a todas las IA, no solo a las IA de uso general con riesgo sistémico y a la IA de alto riesgo. Ya afirmaba MUÑOZ VELA[119] que esto "evidencia una omisión de un requerimiento ético y jurídico que debería ser esencial para cualquier sistema inteligente (por ejemplo un chatbot), sea considerado de nivel medio o bajo o simplemente todavía no clasificado" y por tanto, a cualquier IA generativa ya que el acceso a los datos "de la más inocente" de las IA puede utilizarse para generar confusión, podría afectar a la privacidad

118 También se alude a la ciberseguridad en la IA de alto riesgo, entre otros, en los Considerandos 54,55,74,76,77,78,122,131, art.13.3b) ii, Anexo IV punto 2 h) y art.42.2 LIA. También se alude a la ciberseguridad en los arts.31.2, 58.2 i), 66 h),70.3 y 78.2 LIA

119 MUÑOZ VELA, J.M, *Retos*...op.cit.pág.105

de los datos e incluso provocar riesgos mayores, por ello, todas necesitan un control mínimo. Es así reclamable la incorporación de unas normas mínimas de ciberseguridad para todo tipo de IA y evidentemente, ciberseguridad que puede y debe ser intensificada en los casos de alto riesgo y de riesgo sistémico. Las obligaciones de seguridad y concretamente de ciberseguridad están conectadas con la protección de los datos.

Tal y como expone la LIA en su Considerando 76: "La ciberseguridad es fundamental para garantizar que los sistemas de IA resistan a las actuaciones de terceros maliciosos que, aprovechando las vulnerabilidades del sistema, traten de alterar su uso, comportamiento o funcionamiento o de poner en peligro sus propiedades de seguridad". Añade que los ciberataques contra sistemas de IA pueden dirigirse contra activos específicos de la IA, como los conjuntos de datos de entrenamiento o los modelos entrenados, o aprovechar las vulnerabilidades de los activos digitales del sistema de IA o la infraestructura de TIC subyacente. Aunque el Considerando (y el art.15 LIA) se refiere a la IA de alto riesgo[120], esa afirmación podría aplicarse a cualquier IA, pues puede ser objeto de estos ciberataques expuestos.

Las obligaciones de transparencia

la IA generativa en su camino hacia la perfección de sus capacidades, va mejorando los "errores" y si, por ejemplo, en los vídeos sintéticos iniciales se podía percibir que estábamos en presencia de una inteligencia artificial, por ejemplo, en la mirada vacía del "efecto del valle inquietante", ahora es difícil de descubrir. La misma dificultad está al identificar si la voz sintética es humana

120 Añadía: "Por lo tanto, para garantizar un nivel de ciberseguridad adecuado a los riesgos, los proveedores de sistemas de IA de alto riesgo deben adoptar medidas adecuadas, como los controles de seguridad, teniendo también en cuenta, cuando proceda, la infraestructura de TIC subyacente".

o artificial[121]. En su afán por acercarse, lo más posible al ser humano, aumentan las capacidades de la IA y con ello la posible dificultad de diferenciar lo que es real de lo que no lo es. Actualmente, ChatGPT ha realizado progresos con su último modelo ChatGPT40 que incorpora, además de múltiples funciones, mejoras conversacionales (voces cada vez más "humanas") y la reproducción de sentimientos. El sistema no está reconociendo emociones, sino que intenta reproducirlas, buscando superar uno de los escollos que le separaba del ser humano, al decir que las máquinas no tienen sentimientos, cosa que ellas mismas afirman cuando se les pregunta por algo relacionado con estos temas. Pese a todas las mejoras, aún hay fallas en el sistema, por ejemplo, en la demostración la IA confundió al presentador sonriente con una superficie de madera y también, comenzó a resolver una ecuación que aún no se le había mostrado, por ello aún queda camino que andar, hay fallos, alucinaciones que convierten a los chatbots en potencialmente inseguros y poco fiables[122]. Lo que es innegable es que cada vez va a ser más difícil diferenciar que es real y que es artificial, contenido generado o manipulado con la IA generativa, siendo necesaria la identificación del origen de la creación.

Por otra parte, el riesgo de la oscuridad de la IA en general, y concretamente de la generativa, conlleva una obligación de transparencia, cuyo cumplimiento en principio no conlleva la licitud del uso[123]. Las obligaciones de transparencia de los proveedores y

121 FRANGANILLO, J. "La inteligencia artificial...op. cit. págs.8 y 9. Siguiendo a NIGHTINGALE, S.J, y FARID, H., "Los rostros sintetizados por IA son indistinguibles de los rostros reales y más confiables" Proc Natl Acad Sci US A. 2022 22 de febrero; 119(8): e2120481119. Publicado en línea el 14 de febrero de 2022. doi: 10.1073/pnas.2120481119 (recuperado el 28 de mayo).

122 Tal y como muestra el artículo de KLEINMAN, Z., El tiempo, novedades tecnológicas, "Las seis nuevas funciones de la última versión de ChatGPT: es capaz de coquetear y detectar emociones" ,15 de mayo del 2024, (https://www.eltiempo.com/tecnosfera/novedades-tecnologia/6-nuevas-funciones-de-la-ultima-version-de-chatgpt-que-es-capaz-de-coquetear-y-detectar-emociones-y-las-fallas-que-cometio-3342771) (recuperado el 30 de mayo del 2024).

123 Así expone el Considerando 137 LIA que "el cumplimiento de las obligaciones de transparencia aplicables a los sistemas de IA que entran en el ámbito de aplicación del

responsables de sistemas de IA generativa se reflejan en capítulo IV de la LIA que recoge las obligaciones de transparencia de los proveedores y responsables del despliegue de determinados sistemas de IA.

La exigencia de transparencia fundamental es que las personas que interactúan directamente con una IA estén informadas de que están actuando con un sistema de IA y no con un ser humano, así se dispone que "1. Los proveedores garantizarán que los sistemas de IA destinados a interactuar directamente con personas físicas se diseñen y desarrollen de forma que las personas físicas de que se trate estén informadas de que están interactuando con un sistema de IA" (art.50 LIA).

Si bien establece una excepción "cuando resulte evidente desde el punto de vista de una persona física razonablemente informada, atenta y perspicaz, teniendo en cuenta las circunstancias y el contexto de utilización". Pienso que realmente, la excepción supone una carga excesiva sobre el usuario, la persona que interactúa con la IA, que podría evitarse obligando en todo momento a la identificación o, en su caso, con carácter excepcional a la persona que por sus conocimientos profesionales deberían reconocerlo (Lex artis). Lógicamente, cuando la IA adquiere la forma de un avatar que es claramente identificable y no es idéntico o prácticamente idéntico a una figura humana sería fácil diferenciarlo, con lo que no se genera confusión, pero ¿hasta qué punto es evidente en otras situaciones y se puede exigir el conocimiento de la interacción con la IA? ¿qué grado de diligencia sería exigible?

presente Reglamento no debe interpretarse como un indicador de que la utilización del sistema de IA o de sus resultados de salida es lícito en virtud del presente Reglamento o de otras disposiciones del Derecho de la Unión y de los Estados miembros, y debe entenderse sin perjuicio de otras obligaciones de transparencia aplicables a los responsables del despliegue de sistemas de IA establecidas en el Derecho de la Unión o nacional".

En la misma línea, de considerar la excepción antes expuesta (y propuesta) a los profesionales, está la siguiente excepción legal[124] que excluye de la obligación cuando son los profesionales los que utilizan el sistema respecto a delitos, con la excepción de que el sistema esté a disposición del público para denunciar el delito penal, ya que implica una interactuación con persona física que puede desconocer que está interactuando con la IA.

El segundo punto del art.50 LIA, directamente se refiere a la IA generativa, al afirmar "los proveedores de sistemas de IA, entre los que se incluyen los sistemas de IA de uso general, que generen contenido sintético de audio, imagen, vídeo o texto, velarán por que los resultados de salida del sistema de IA estén marcados en un formato legible por máquina y que sea posible detectar que han sido generados o manipulados de manera artificial". Se centra en cómo tienen que ser las soluciones técnicas[125] y excluye de la obligación a "los sistemas de IA que desempeñen una función de apoyo a la edición estándar o no alteren sustancialmente los datos de entrada facilitados por el responsable del despliegue o su semántica, o cuando estén autorizados por ley para detectar, prevenir, investigar o enjuiciar delitos".

En su punto cuarto, el artículo distingue entre los responsables del despliegue de un sistema de IA que genere o manipule imágenes o contenidos de audio o vídeo que constituyan una ultrasuplantación[126] y, por otra parte, texto que se publique con el fin de informar al público sobre asuntos de interés público.

124 "Esta obligación no se aplicará a los sistemas de IA autorizados por ley para detectar, prevenir, investigar o enjuiciar delitos, con sujeción a las garantías adecuadas para los derechos y libertades de terceros, salvo que estos sistemas estén a disposición del público para denunciar un delito penal".

125 "Los proveedores velarán por que sus soluciones técnicas sean eficaces, interoperables, sólidas y fiables en la medida en que sea técnicamente viable, teniendo en cuenta las particularidades y limitaciones de los diversos tipos de contenido, los costes de aplicación y el estado actual de la técnica generalmente reconocido, según se refleje en las normas técnicas pertinentes".

126 Art.3 concepto 60 "ultrasuplantación: un contenido de imagen, audio o vídeo generado o manipulado por una IA que se asemeja a personas, objetos, lugares, entidades

En el primer caso (imágenes, audio y video) de ultrasuplantación, "harán público que estos contenidos o imágenes han sido generados o manipulados de manera artificial. Esta obligación no se aplicará cuando la ley autorice su uso para detectar, prevenir, investigar o enjuiciar delitos". Esta obligación de transparencia aparece matizada, en su Considerando 134, nos muestra que esta obligación además de con las soluciones técnicas utilizadas, se consigue a través del etiquetado de los resultados de salida generados por la IA. Se aclara en el artículo, que cuando el contenido forme parte de una obra o programa manifiestamente creativos, satíricos, artísticos, de ficción o análogos, estas obligaciones de transparencia se limitarán "a la obligación de hacer pública la existencia de dicho contenido generado o manipulado artificialmente de una manera adecuada que no dificulte la exhibición o el disfrute de la obra".

En relación al punto cuarto, sobre la intervención de la IA generativa en una creación artística, señalar que los artistas cada vez más están usando las redes neuronales para sus obras. Esto no impide la valoración de la misma en sí, su apreciación, pero siempre partiendo de la base de que somos conocedores del origen o de la intervención de la IA generativa en su creación. Ha de recordarse que hay obras sintéticas que han sido premiadas desconociendo su origen. Debe respetarse el derecho a la creación artística. La obligación de transparencia, que nos libre de la confusión y la desinformación, el saber que la obra ha sido generada o manipulada artificialmente, debe buscar el camino adecuado para no afectar a la obra en sí y la posibilidad de su disfrute. El Ministerio de Cultura ha dejado claro que las obras exclusivamente generadas con IA no podrán ser premiadas[127].

o sucesos reales y que puede inducir a una persona a pensar erróneamente que son auténticos o verídicos".

127 "Las obras creadas "exclusivamente" con Inteligencia Artificial no podrán ganar un Premio Nacional de Cultura" El Periódico de España, Cultura, noticias efe, 19 de febrero de 2024 (https://www.epe.es/es/cultura/20240219/obras-creadas-exclusivamente-inteligencia-artificial-codigo-buenas-practicas-ministerio-cultura-98374508) (recuperado el 13 de mayo

En el segundo caso expuesto en la norma, si se refiere a texto que se publique con el fin de informar al público sobre asuntos de interés público, los responsables "divulgarán que el texto se ha generado o manipulado de manera artificial. Esta obligación no se aplicará cuando el uso esté autorizado por ley para detectar, prevenir, investigar o enjuiciar delitos, o cuando el contenido generado por IA haya sido sometido a un proceso de revisión humana o de control editorial y cuando una persona física o jurídica tenga la responsabilidad editorial por la publicación del contenido"[128]. Se prevé una obligación de divulgación similar a la anterior. Debe recordarse que la generación automática de textos etc. sin control puede provocar la desinformación. Es por ello necesario el control humano, pero ese control humano no debe excluir en ningún caso la identificación de la actuación de la Inteligencia Artificial, en nuestro caso generativa. Aunque tengamos el control de terceras personas o la responsabilidad reconocida a una persona por la publicación del contenido, hay que actuar con prevención no asegurando sin más la responsabilidad concreta "humana" sino evitando el problema (que se produzca el daño) con la referencia, en todo caso, a la intervención de la IA en el texto.

Esta obligación de transparencia del uso de la IA generativa o de sus resultados (en ambos supuestos) no obstaculiza el derecho a la libertad de expresión ni al de creatividad artística y científica, garantizados por la Carta de los Derechos Fundamentales de

del 2024). El Ministerio de Cultura ha elaborado una guía de buenas prácticas relativas a la IA para regular su uso y se aplicarán en: la contratación de actividades y servicios creativos, los Premios Nacionales y las subvenciones.

128 En esta línea de control, el Reglamento (UE) 2024/1083 del Parlamento Europeo y del Consejo, de 11 de abril de 2024, por el que se establece un marco común para los servicios de medios de comunicación en el mercado interior y se modifica la Directiva 2010/13/UE (Reglamento Europeo sobre la Libertad de los Medios de Comunicación), DOUE, nº 1083, de 17 de abril de 2024 establece en su art.18.1 que "Los prestadores de plataformas en línea de muy gran tamaño facilitarán una funcionalidad que permita a los destinatarios de sus servicios: e) declarar que no ofrecen contenidos generados por sistemas de inteligencia artificial sin someterlos a revisión humana o control editorial".

la Unión Europea (arts.11 y 13), particularmente cuando el contenido forme parte de una obra o programa manifiestamente creativos, satíricos, artísticos, de ficción o análogos, con sujeción a unas garantías adecuadas para los derechos y libertades de terceros, como expone el Considerando 134.

En el apartado 5 del art.50 LIA se afirma que la información de los apartados anteriores "se facilitará a las personas físicas de que se trate de manera clara y distinguible a más tardar con ocasión de la primera interacción o exposición". La inmediatez del aviso (que estamos ante una actuación de la IA) evita el problema, desde el primer acercamiento a la IA y debe conocerse[129]. Como muestra el artículo, desde el punto de vista formal la información debe ser clara y distinguible y ofrecerse, como muy tarde desde la primera interacción o exposición. Añade el artículo que "la información se ajustará a los requisitos de accesibilidad aplicables"[130].

En relación a la obligación de detectar y divulgar que los resultados son originados por la IA generativa, en el Considerando 136, se expresa "que las obligaciones impuestas a los proveedores y a los responsables del despliegue de determinados sistemas de IA en el presente Reglamento destinadas a permitir que se detecte y divulgue que los resultados de salida de dichos sistemas han sido generados o manipulados de manera artificial resultan especialmente pertinentes para facilitar la aplicación efectiva del Reglamento (UE) 2022/2065" del Parlamento Europeo y del Consejo de 19 de octubre de 2022 relativo a un mercado único de servicios digitales y por el que se modifica la Directiva 2000/31/CE (Reglamento de Servicios Digitales).

129 Entiendo que al hacer referencia al primer acercamiento carece de sentido "a más tardar" con la ocasión de la primera interacción o exposición, podría ser más conveniente "desde" la primera interacción o exposición.

130 Manifiesta el art.50 en su punto 6 que, estas obligaciones (de los apartados 1 a 4) no afectarán a los requisitos y obligaciones establecidos en el capítulo III de la LIA y "se entenderán sin perjuicio de otras obligaciones de transparencia establecidas en el Derecho nacional o el de la Unión para los responsables del despliegue de sistemas de IA".

Se aplica particularmente, según el mencionado Considerando, "en lo referente a las obligaciones de los prestadores de plataformas en línea de muy gran tamaño o de motores de búsqueda en línea de muy gran tamaño para detectar y mitigar los riesgos sistémicos que pueden surgir de la divulgación de contenidos que hayan sido generados o manipulados de manera artificial, en particular el riesgo de los efectos negativos reales o previsibles sobre los procesos democráticos, el discurso cívico y los procesos electorales, como a través de la desinformación. La exigencia de etiquetar los contenidos generados por sistemas de IA con arreglo al presente Reglamento se entiende sin perjuicio de la obligación prevista en el artículo 16, apartado 6, del mencionado Reglamento (UE) 2022/2065[131]".

Un claro ejemplo de la información a las personas físicas del uso de la IA generativa y su detección y etiquetado lo tenemos en Instagram (Meta). Ahora cuando se sube una imagen a las historias, se hace una referencia a que se etiquete el uso de IA (pone etiquetar IA) y establece unas normas sobre el etiquetado de contenido generado con IA en Instagram. Se pide una actuación por parte del usuario y se reconoce la actividad de detección que realiza Instagram (explica por qué debes etiquetar el contenido generado con IA en Instagram, cómo funciona el etiquetado de IA en Instagram, cuándo es obligatoria la etiqueta "creado con IA"[132] y cómo etiquetar contenido generado con IA).

131 Añade, "para los prestadores de servicios de alojamiento de datos de tratar las notificaciones que reciban sobre contenidos ilícitos en virtud del artículo 16, apartado 1, de dicho Reglamento, y no debe influir en la evaluación y la decisión sobre el carácter ilícito del contenido de que se trate. Dicha evaluación debe realizarse únicamente con referencia a las normas que rigen la legalidad del contenido".

132 "Meta exige que etiquetes el contenido que compartas y que contenga vídeos fotorrealistas o audios que parezcan realistas generados o alterados digitalmente, incluso con IA. Esto significa que, si este tipo de contenido se crea o se modifica con una herramienta de IA o creación digital, debes etiquetarlo antes de compartirlo. Meta no te exige que etiquetes imágenes que se hayan creado o alterado con IA. Sin embargo, estas imágenes recibirán la etiqueta si nuestros sistemas detectan que se generaron o modificaron mediante IA. Nota: Podría haber sanciones si no etiquetas el contenido según proceda. Aquí te mostramos algunos ejemplos de contenido creado digitalmente que debe etiquetarse: Un vídeo de un

"La oficina de IA fomentará y facilitará la elaboración de códigos de buenas prácticas a escala de la Unión para promover la aplicación efectiva de las obligaciones relativas a la detección y el etiquetado de contenidos generados o manipulados de manera artificial" (apartado 7 art.50 LIA)[133]. En el cumplimiento de las obligaciones de transparencia juegan un papel esencial los códigos de buenas prácticas. "Sin perjuicio del carácter obligatorio y de la plena aplicabilidad de las obligaciones de transparencia", señala el Considerando 135, "la Comisión podrá también fomentar y facilitar la elaboración de códigos de buenas prácticas a escala de la Unión, a fin de facilitar la aplicación eficaz de las obligaciones en materia de detección y etiquetado" mencionadas. Añade el Considerando que "también para apoyar disposiciones prácticas para que, según proceda, los mecanismos de detección sean accesibles y facilitar la cooperación con otros agentes de la cadena de valor, difundiendo los contenidos o comprobando su autenticidad y procedencia, a fin de que el público pueda distinguir efectivamente los contenidos generados por IA"[134].

grupo de personas caminando por un mercado al aire libre que parezca realista, Un archivo de audio de dos personas hablando, una canción creada por voces generadas por IA, un reel narrado con una voz en off realista generada con IA. Aquí te mostramos algunos ejemplos de contenido creado digitalmente que no es necesario que etiquetes: un vídeo de un paisaje al aire libre creado con un estilo similar al de los dibujos animados, un vídeo en el que se ha cambiado ligeramente el tamaño y se ha recortado mínimamente. Puedes obtener más información sobre este requisito en el Centro de transparencia" (Información obtenida en Instagram).

133 Se añade que "La Comisión podrá adoptar actos de ejecución a fin de aprobar dichos códigos de buenas prácticas, de conformidad con el procedimiento establecido en el artículo 56, apartado 6. Si considera que el código no es adecuado, la Comisión podrá adaptar un acto de ejecución que especifique normas comunes para el cumplimiento de las citadas obligaciones de conformidad con el procedimiento de examen establecido en el artículo 98, apartado 2". El Considerando 117 trata sobre la importancia de los códigos de buenas costumbres en el cumplimiento de las obligaciones de los proveedores de modelos de IA de uso general.

134 Por su parte, el Considerando 107 LIA establece para aumentar la transparencia en relación con los datos utilizados en el entrenamiento previo y el de los modelos de IA de uso general, respetando los derechos de autor, que los proveedores de esos modelos elaborarán y pondrán a disposición del público un resumen detallado de los contenidos usados para el entrenamiento.

La ética

La LIA busca una IA fiable, de confianza, afirma en su Considerando 27 que el "enfoque basado en el riesgo es la base de un conjunto proporcionado y eficaz de normas vinculantes", reconociéndose la importancia de recordar las Directrices éticas para una IA fiable. La fiabilidad nos la da el conocimiento, la alfabetización, las medidas de transparencia e información, entre otras cosas, y principalmente las Directrices éticas. En esta línea, menciona las Directrices éticas para una IA fiable, de 2019, elaboradas por el Grupo Independiente de Expertos de Alto Nivel sobre IA creado por la Comisión, en las que se desarrollan siete principios éticos no vinculantes para la IA que tienen por objeto contribuir a garantizar la fiabilidad y el fundamento ético de la IA. Dichos principios son: acción y supervisión humanas; solidez técnica y seguridad; gestión de la privacidad y de los datos; transparencia; diversidad, no discriminación y equidad; bienestar social y ambiental, y rendición de cuentas. Continúa reconociendo que "sin perjuicio de los requisitos jurídicamente vinculantes del presente Reglamento y de cualquier otro acto aplicable del Derecho de la Unión, esas directrices contribuyen al diseño de una IA coherente, fiable y centrada en el ser humano, en consonancia con la Carta y con los valores en los que se fundamenta la Unión". De acuerdo con las directrices del Grupo de Expertos, "por "acción y supervisión humanas" se entiende que los sistemas de IA se desarrollan y utilizan como herramienta al servicio de las personas, que respeta la dignidad humana y la autonomía personal, y que funciona de manera que pueda ser controlada y vigilada adecuadamente por seres humanos".

En el texto se incorporan así, expresamente, las Directrices éticas básicas (acción y supervisión humanas; solidez técnica y seguridad; gestión de la privacidad y de los datos; transparencia; diversidad, no discriminación y equidad; bienestar social y ambiental, y rendición de cuentas) pero se pierde la oportunidad de convertirlas directamente en deberes jurídicos exigibles a todo tipo

de IA, además teniendo en cuenta que ya fueron incorporadas como tales y desapareció posteriormente el artículo que las contemplaba (4 bis)[135]. Esto sería, independientemente de que estas directrices sean la base de la regulación del Reglamento, con las correspondientes matizaciones o añadidos en función del mayor o menor riesgo del sistema de IA. Se exigirían, entonces a las IA generativas (incide el Considerando en el carácter no vinculante de las mismas). Destacar, la acción y supervisión humanas, lo que aleja de las IA autónomas que tengan la última decisión. Quedan así, reflejadas expresamente en dicho Considerando, sin incorporarlas expresamente como normas de carácter general aplicable a todas las IA en el articulado del Reglamento[136].

Convertir las exigencias éticas en jurídicas marcaría las diferencias con otros países, facilitaría el punto de partida de actuación en materia de IA con ellos. Mostraría a la ciudadanía el auténtico valor del tratamiento ético, e iría calando en la misma. En la reciente Convención Marco del Consejo de Europa sobre Inteligencia Artificial y Derechos Humanos, Democracia y Estado de Derecho de 17 de mayo del 2024[137], se ha vuelto a perder otra

135 Fueron considerados principios generales aplicables a todos los sistemas de IA en las enmiendas aprobadas por el Parlamento Europeo el 14 de junio de 2023 en su Enmienda 213 en la que se incluyó una propuesta de Reglamento con un artículo 4 bis que los recogía normativamente como principios generales aplicables a todos los sistemas de IA (Enmiendas aprobadas por el Parlamento Europeo el 14 de junio de 2023 sobre la propuesta de Reglamento del Parlamento Europeo y del Consejo por el que se establecen normas armonizadas en materia de inteligencia artificial (Ley de Inteligencia Artificial) y se modifican determinados actos legislativos de la Unión (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)), Procedimiento legislativo ordinario: primera lectura , Enmienda 213 , Propuesta de Reglamento, artículo 4 bis (nuevo)). Señalando en su apartado 1 que " Todos los operadores que entren en el ámbito de aplicación del presente Reglamento se esforzarán al máximo por desarrollar y utilizar los sistemas de IA o modelos fundacionales con arreglo a los siguientes principios generales que establecen un marco de alto nivel para promover un enfoque europeo coherente centrado en el ser humano con respecto a una inteligencia artificial ética y fiable, que esté plenamente en consonancia con la Carta, así como con los valores en los que se fundamenta la Unión".

136 MUÑOZ VELA, J.M, *Retos*...op.cit.págs.100 y 101, considera un riesgo la ausencia de normas éticas vinculantes , estando a favor de las mismas.

137 13ª Sesión del Comité de Ministros (Estrasburgo, 17 de mayo de 2024), Comisión de Inteligencia Artificial (CAI), CM (2024)52-final.

oportunidad de dar valor jurídico exigible a las exigencias éticas esenciales en el modo de entender cualquier tipo de IA , recogiéndolas expresamente en el articulado como normas éticas.

No queda más que afirmar que la IA generativa, al servicio del hombre, debe seguir la senda de la constante evolución, guiada de mano humana sustentada en una ética exigible, para que el camino de rosas no se convierta en camino de espinas, para que Gizmo no genere seres como Stripe que puedan poner en jaque a la humanidad.

V.- CONCLUSIONES

1. La LIA no define directamente la IA generativa, la engloba en la IA de uso general. La transformación en la forma de generar contenidos de la IA generativa nos proporciona múltiples beneficios en todas las esferas de nuestra vida. La inteligencia artificial generativa permite la creación automática de contenido "original" en diferentes formatos (texto, audio, vídeo e imágenes). Un contenido difícil de diferenciar del creado por humanos. Ofrece tantos beneficios como posibles riesgos para la sociedad, democracia y derechos fundamentales. Son contenidos tan "reales" que pueden llevar a la confusión y a la desinformación, desconociendo que es real y que no. Provocan la manipulación y actuaciones delictivas a través, principalmente de fake news, deepvoice y deepfakes. Estos últimos son especialmente peligrosos, los vídeos hipertrucados, las ultrasuplantaciones, siendo una de las mayores preocupaciones de los gobiernos.
2. Son numerosos los riesgos que puede conllevar la IA generativa. Los riesgos de desinformación, confusión, ataque al derecho al honor, la intimidad y la propia imagen (provocados

principalmente por los deepfakes) que tiene que convivir con el derecho a la libertad de información y expresión. El riesgo de infracción y vulneración de los derechos de autor, ante creaciones que se nutren de otras de acceso público por internet sin retribución inicial y con posibles problemas de plagio y copyright, en las que serán necesarios los correspondientes consentimientos. El riesgo en el mercado laboral, con la supresión de puestos de trabajo, si bien aparecen nuevos trabajos y cualificación laboral e incremento de la productividad. Los riesgos del medio ambiente ante el gasto energético que supone y la busca de salidas como la energía nuclear. Los de sesgos al poder reproducir estereotipos (y producirse discriminaciones) con datos sin filtrar, cuyo entrenamiento tiene que calibrarse y reentrenarse, e incluso minimizarse el uso de datos. Es necesaria la alfabetización, es decir la comprensión y capacitación digital en materia de IA generativa y la democratización en su acceso y comprensión. Riesgos por su uso malicioso, lo que va a exigir un control para mitigar los efectos negativos y la desinformación generada por la confusión o por modelos mal entrenados. Riesgo de toxicidad al reproducir el lenguaje de la red. Riesgos por errores y alucinaciones, ya que la IA generativa crea respuestas lógicas que pueden ser falsas. Riesgo para la seguridad, principalmente de los datos personales, ante posibles fugas o uso indebido o datos inexactos, y además el riesgo para la privacidad, de la memorización y réplica posterior de datos por la IA y el posible peligro para la confidencialidad. La ciberseguridad es esencial ante ataques como, por ejemplo, el del "fraude del CEO". Riesgos psicológicos ante fenómenos como "revivir" a personas fallecidas, la posible exclusión laboral o sextorsion. Riesgo en relación a cuestiones éticas y la exigibilidad de Códigos de Conducta, y en relación a la formulación de las instrucciones en la alienación de valores.

3. La IA generativa podría incluirse directamente en la IA prohibida en los supuestos 5.1 a) y 5.1b) LIA. En el momento que das una información, imagen o vídeo falsos o manipulados (deepfakes, fake news, deep voice) que no permite distinguir fácilmente la realidad de la ficción (sobre todo con los deepfakes), de forma subliminal o deliberada se puede manipular y engañar, con el objetivo o efecto de alterar de manera sustancial, el comportamiento de una persona o colectivo. Lo que impide tomar una decisión informada y le o les lleve a tomar otra que le o les provoca o puede provocar perjuicios considerables. Si bien, interpreto en el caso de hacerse de forma subliminal que debe estar prohibida siempre. Igual ocurre cuando se exploten vulnerabilidades para alterar sustancialmente el comportamiento, haciendo creer lo que no es, y estas personas no son capaces de discernir, creando confusión, desinformación y manipulación. En los otros supuestos de Inteligencias Artificiales prohibidas, la prohibición sería indirecta, podría la IA generativa servirles, por ejemplo, creándoles datos sintéticos, en estos casos su uso, también debería estar prohibido.
4. La IA generativa puede ser IA general con riesgos sistémicos. Además de la presencia de posibles capacidades de gran impacto, es indudable que la IA generativa puede tener efectos negativos reales o razonablemente previsibles en la salud pública, la seguridad, la seguridad pública, los derechos fundamentales o la sociedad en su conjunto. Queda corroborado al analizar previamente sus riesgos. Estos riesgos pueden propagarse a gran escala a lo largo de toda la cadena de valor. A su vez, queda reflejada esa posibilidad en el Considerando 136 de la LIA. La IA generativa podría incluirse directamente en el supuesto de IA del alto riesgo en los Sistemas de IA destinados a influir en el resultado de una elección o referéndum, por lo antes expuesto (la posible manipulación, confusión y desinformación). Debería incluir-

se directamente el deepfake como IA de alto riesgo, al poner en riesgo numerosos derechos fundamentales y poder generar confusión y desinformación que afecte a la capacidad de decisión en una elección o referéndum (y en cualquier decisión). En esta línea está el Estudio, la política europea frente a los deepfakes, de julio del 2021.

5. Se hace especial referencia a las obligaciones de ciberseguridad y transparencia. La ciberseguridad es tan importante que, pese a abrirse el abanico de su exigencia además de a la IA de alto riesgo (art.15 LIA) a la de uso general con riesgo sistémico (art.55 .1 d) LIA), a diferencia de redacciones anteriores de la LIA (2021), realmente debería ser una exigencia para todas las Inteligencias Artificiales. La IA "más inocente" puede abrir la puerta a la confusión, afectar a la privacidad de datos y provocar riesgos mayores. Obligación de transparencia, ante la oscuridad de la IA generativa, reflejada en el art.50 LIA. Es fundamental que las personas físicas sepan que están interactuando con una inteligencia artificial siendo criticable su excepción, en relación al posible conocimiento por parte de la persona física. Un aspecto destacable es la importancia de la ética. La LIA recoge en su Considerando 27,7 Directrices éticas básicas (acción y supervisión humanas; solidez técnica y seguridad, gestión de la privacidad y de los datos, transparencia, diversidad, no discriminación y equidad, bienestar social y ambiental y rendición de cuentas) perdiendo la oportunidad de convertirlas en jurídicas, en normas comunes a todo tipo de IA (incluida la generativa), independientemente de que queden reflejadas en su articulado.

VI.- BIBLIOGRAFÍA

ABDULKABBER, T., "Funcionarios y famosos: así circuló la información falsa en las elecciones nigerianas", ijnet (red internacional de periodistas), Lucha contra la desinformación, 13 de abril del 2023 ,https://ijnet.org/es/story/funcionarios-y-famosos-as%C3%AD-circul%C3%B3-la-informaci%C3%B3n-falsa-en-las-elecciones-nigerianas) (recuperado el 3 de agosto del 2024).

AFP, "La fiscalía investiga si Meta vulnera la protección de datos de sus usuarios", El Mundo, Empresas, 4 de julio de 2024 (recuperado el 7 de julio del 2024).

ÁLVAREZ, P. y EGUILUZ, J. "El Reglamento de IA ante los deepfakes de desnudos", 2 de octubre del 2023, Cuatrecasas (https://www.cuatrecasas.com/es/spain/propiedad-intelectual/art/el-reglamento-de-ia-ante-los-deepfakes-de-desnudos) (recuperado el 1 de mayo del 2024).

ANDRÉS, R., "La última tendencia para bordar las entrevistas de trabajo: entrenar con ChatGPT como reclutador", Xataca, 6 de marzo del 2024 (recuperado el 28 de julio del 2024).

ANNEMANS,R.,"Seguridad de la IA generativa: 8 riesgos que debes conocer". GlobalSing by GMO, 4 de diciembre del 2023 https://www.globalsign.com/es/blog/8-riesgos-de-seguridad-de-la-inteligencia-artificial-generativa (recuperado el 10 de mayo del 2024).

BANDARA, P., "Elon Musk compartió un vídeo engañoso de inteligencia artificial de la vicepresidenta Kamala Harris en X", PetaPixel, 29 de julio del 2024 (https://petapixel-com.translate.goog/2024/07/29/elon-musk-shared-misleading-ai-video-of-kamala-harris-on-x-twitter-deepfake-election-presidential-usa/?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=sc) (recuperado el 3 de agosto del 2024).

BARRIO ANDRÉS, M., "Algunos claroscuros en el Reglamento Europeo de Inteligencia Artificial", *Diario La Ley*, nº 86. Sección Ciberderecho, 30 de julio de 2024.

BENDITO CAÑIZARES, M.T, "Estadio intermedio de reflexión para una futura regulación de la ética en el espacio digital europeo: los principios de transparencia y accountability", *Revista Aranzadi de Derecho y Nuevas Tecnologías*, núm 55/2021, BIB 2021/1465.

BORRACHERO GARRO, A. "Los retos de la tecnología en la publicidad: la campaña de Lola Flores", coordinador: ORTEGA BURGOS, E., *Propiedad intelectual*, 2022, Documento TOL9.141.406.

CALDERÓN C.,"Elecciones México 2024: Deep fakes y fake news ganan en las votaciones", El financiero, 4 de junio del 2024,https://www.elfinanciero.com.mx/elecciones-mexico-2024/2024/06/04/deep-fakes-y-fake-news-marcaron-el-escenario-electoral/ (recuperado el 3 de agosto del 2024).

CASTILLO RAMOS-BOSSINI, S.E., "Regulación europea de la inteligencia artificial", *Nuevas fórmulas de prestación de servicios en la era digital*, dirección Juan Francisco Pérez Gálvez, Dyinson, 2023.

DAVID, DemoCreator, "Los 10 mejores generadores de música IA gratis en 2024", 13 de marzo de 2024 (https://dc.wondershare.es/ai-voice/top-free-ai-music-generators.html) (recuperado el 7 de mayo del 2024).

DEL CASTILLO, C., "Los creadores del canon AEDE quieren una "tasa ChatGPT" para la inteligencia artificial, el Diario.es, 3 de mayo del 2023, (https://www.eldiario.es/tecnologia/creadores-canon-aede-quieren-tasa-chatgpt-inteligencia-artificial_1_10171676.html), (recuperado el 20 de mayo del 2024).

DURAN, I., "Taylor Swift y sus desnudos hechos con IA: CEO de Microsoft dice "hay que actuar ya" ante los deepfakes, In-

fobae, 29 de enero del 2024, https://www.infobae.com/tecno/2024/01/27/taylor-swift-y-sus-desnudos-hechos-ia-ceo-de-microsoft-dice-hay-que-actuar-ya-ante-los-deepfakes/ (rescatado el 3 de agosto del 2024).

ESPUGA TORNÉ, G. "Cómo identificar contenido generado por IA" Linkedin, junio, 2024 (recuperado el 10 de julio del 2024).

FIGUEROA, J.C., "Crear una sola imagen con inteligencia artificial consume tanta energía como cargar tu teléfono", Hipertextual, Tecnología, 12 de diciembre de 2023 (https://hipertextual.com/2023/12/crear-imagen-con-inteligencia-artificial-consume-esta-energia) (recuperado el 20 de abril del 2024).

FERNÁNDEZ HERNÁNDEZ, C. y EGUILUZ CASTAÑEIRA, J., "Diez puntos críticos del Reglamento europeo de Inteligencia Artificial", *Diario LA LEY*, Sección Ciberderecho, nº 85, 28 de junio de 2024.

FRANGANILLO, J., "La inteligencia artificial generativa y su impacto en la creación de contenidos mediáticos", *methaodos. revista de ciencias sociales* (2023) 11(2) m231102a1010.17502/mrcs.v11i2.710.

GAMERO CASADO, E. "El enfoque europeo de la Inteligencia Artificial", *Revista de Derecho administrativo*, nº 20, 2021, págs.268 y ss.

GARAY, J., Wired, 19 de mayo del 2023, "El G 7 promete regular las IA generativas antes de que termine el 2023"(https://es.wired.com/articulos/g7-promete-regular-las-ia-generativas-antes-de-que-termine-el-2023).

GARCÍA MEXÍA, P. "Europa ante el reto de la inteligencia artificial",-The objective, 3 de agosto del 2024, (https://theobjective.com/tecnologia/2024-08-03/europa-ante-el-reto-de-la-inteligencia-artificial/) (recuperado el 5 de agosto del 2024).

GOLDMAN SACHS, Principales riesgos que entraña la inteligencia artificial generativa: enumeración, fundspeople, (https://fundspeople.com/es/principales-riesgos-que-entrana-la-inteligencia-artificial-generativa/), 26 de abril de 2023 (rescatada en 20 de abril del 2024).

HOLCOMBE, J., "Las 9 Mejores Herramientas de Detección de Contenidos con IA que Tienes que Conocer", Kinsta, 5 de abril del 2023 https://kinsta.com/es/blog/deteccion-de-contenidos-ia/ (recuperada el 6 de julio del 2024).

JARQUES, A., "El futuro Reglamento Europeo de Inteligencia Artificial", *Actualidad Jurídica Aranzadi,* nº1003, 2023, Editorial Aranzadi, (BIB 2024/278) (rescatada en 27 de mayo del 2024).

KLEINMAN, Z., El tiempo, novedades tecnológicas, "Las seis nuevas funciones de la última versión de ChatGPT: es capaz de coquetear y detectar emociones" ,15 de mayo del 2024, (https://www.eltiempo.com/tecnosfera/novedades-tecnologia/6-nuevas-funciones-de-la-ultima-version-de-chatgpt-que-es-capaz-de-coquetear-y-detectar-emociones-y-las-fallas-que-cometio-3342771) (recuperado el 30 de mayo del 2024).

LALCHAND, S. et al. (Centro de Servicios Financieros de Deloitte), "Se espera que la IA generativa aumente el riesgo de deepfakes y otros fraudes de la banca", Deloitte, Servicios Financieros, 29 de mayo del 2024, https://www2.deloitte.com/us/en/insights/industry/financial-services/financial-services-industry-predictions/2024/deepfake-banking-fraud-risk-on-the-rise.html)) (recuperado el 2 de agosto del 2024).

LINDSAY, J. M., "Elecciones 2024: La amenaza falsa a las elecciones del 2024", Council on Foreign Relations, 2 de febrero, 2024, https://www.cfr.org/blog/election-2024-deepfake-threat-2024-election (recuperado el 3 de agosto del 2024).

LUCIO LÓPEZ, L.A., "Deep fake porn, la inteligencia artificial da nueva cara al ciberacoso escolar", Ciem, 2024.

MARTÍNEZ ESPÍN, P., "La propuesta de marco regulador de los sistemas de inteligencia artificial en el mercado de la UE", *Revista CESCO de Derecho de Consumo*, nº46/2023(doi.org/10.18239/RDC_2023.46.3322).

MCMAHON, L., BBC News, 20 de mayo de 2024, "Cuando la escuché me quedé en shock": por qué el programa de IA ChatGPT dejará de usar la voz que se parece a la de Scarlett Johanson, https://www.bbc.com/mundo/articles/cprrzn8g2wqo.) (recuperado el 30 de mayo del 2024).

MEAKER, M., "Deepfakes en elecciones de Eslovaquia rearfiman que IA es un peligro para la democracia", Wired, negocios, 3 de octubre del 2023 https://es.wired.com/articulos/deepfakes-en-elecciones-de-eslovaquia-reafirman-que-ia-es-peligro-para-democracia) (recuperado el día 2 de agosto del 2024).

MEIJOMIL, S., Inboundcycle, "6 generadores de imágenes con IA que no puedes perderte, 8 de noviembre del 2023, (https://www.inboundcycle.com/blog-de-inbound-marketing/generadores-de-imagenes-con-ia).

MICROSOFT SECURITY, IA generativa, La ventaja de los defensores, Linkedln, Microsoft (recuperado en 6 de mayo de 2024).

MORAN, I., Photolari, 2 de mayo, 2024, "20.000 euros por la imagen IA que engañó al jurado de los Sony Wordl Photography Awards el año pasado" (https://www.photolari.com/20-000-euros-por-la-imagen-ia-que-engano-al-jurado-de-los-sony-world-photography-awards-el-ano-pasado/) (recuperado el 20 de mayo del 2024).

MUÑOZ VELA, J. M, "Inteligencia artificial generativa. Desafíos para la propiedad intelectual", *Revista de Derecho UNED*, núm.33, 2024, Premio de artículos jurídicos "García Goyena", 22ª convocatoria (curso 2022-2023), Facultad de Derecho. UNED.

MUÑOZ VELA, J.M, *Retos, riesgos, responsabilidad y regulación de la inteligencia artificial. Un enfoque de seguridad física, lógica, moral y jurídica*, Thomson Reuters- Aranzadi, Pamplona, 2022.

NIGHTINGALE, S.J, y FARID, H., "Los rostros sintetizados por IA son indistinguibles de los rostros reales y más confiables" Proc Natl Acad Sci US A. 2022 22 de febrero; 119(8): e2120481119. Publicado en línea el 14 de febrero de 2022. doi: 10.1073/pnas.2120481119 (recuperado el 28 de mayo).

PASCUAL, M.G.,Meta no ofrecerá sus nuevos modelos de IA generativa en Europa por su "impredicible entorno regulatorio", El Pais, Tecnología, 18 de julio de 2024 (recuperado el 19 de julio del 2024).

PLAZA PENADÉS, J., Dossier, "Las claves de la futura Ley de Inteligencia Artificial Europea", Aranzadi La Ley, Navarra, mayo, 2023.

PUERTO MENDOZA, A., Derecho *digital. Fundamentos básicos*, Ediciones CEF.,2019.

PUFPAFF, M. "¿La tecnología como fuerza para el bien? ¿Cómo se está utilizando la inteligencia artificial para prevenir los suicidios en China?", *Razón y fe*, Tomo 282, nº1447, 2020, págs.205 y ss.

RAA J., "Meta detiene su proyecto para entrenar a la IA con publicaciones de Facebook e Instagram en Europa", Tecnología, El País, 14 de junio de 2024 (recuperado el 1 de julio del 2024).

SÁNCHEZ, L., "Escrivá advierte que la estrategia de IA necesita de entornos seguros y anuncia una nueva ley integral de ciberseguridad", Economist & Jurist, 7 de junio del 2024 (recuperado en 20 de julio del 2024).

SINCLA, A et al." El estado de la IA a principios de 2024: la adopción de la IA generativa aumenta y comienza a generar valor", McKinsey& Company, 30 de mayo de 2024, (https://www.mckinsey.com/locations/south-america/latam/hispanoamerica-en-potencia/el-estado-de-la-ia-a-principios-de-2024-la-adopcion-de-la-ia-generativa-aumenta-y-comienza-a-generar-valor/es-CL) (recuperado el 2 de agosto de 2024).

SOTO ARAMENDARIZ, S "Primer suicidio inducido por inteligencia artificial: algo que temer", 4 de abril de 2023 (https://observatorioblockchain.com/ia/primer-suicidio-inducido-por-inteligencia-artificial-algo-que-temer) (recuperado el 28 de mayo del 2024).

SUÁREZ JAQUET, H. et HINOJAL CUADRADO, E. "El uso del deepfake en producciones audiovisuales: consideraciones jurídicas", coordinador: ORTEGA BURGOS, E., *Propiedad intelectual*, 2022, Documento TOL9.141.396

TENBARGE, K.,"Taylor Swift, deepfakes en X describe falsamente que apoya a Trump", NBC NEWS, 8 de febrero del 2024, https://www.nbcnews.com/tech/internet/taylor-swift-deepfake-x-falsely-depict-supporting-trump-grammys-flag-rcna137620 (recuperado el 3 de agosto del 2024).

VALERO, A., "Deepfakes Porn y violencia contra las mujeres", Fundación Cañada Blanch, 4 de junio de 2024, https://www.fundacioncanadablanch.org/noticias/deepfakes-porn-y-violencia-contra-las-mujeres/ (recuperado el 2 de agosto del 2024).

VIDAL. M, en Linkedln, noticia, (Fuente: https://www.nytimes.com/2024/05/28/technology/ai-chief-executives.html) (recuperado el 7 de junio de 2024).

VIGARIO, D.,"Un año de libertad vigilada para los 15 jóvenes que manipularon y difundieron imágenes con IA de menores desnudas en Almendralejo", El mundo, 9 de julio de 2024 (recuperado en 1 de agosto del 2024).

WIKIPEDIA, Greemlins, https://es.wikipedia.org/wiki/Gremlins (recuperado el 1 de mayo del 2024).

"Las empresas de inteligencia artificial ofrecen ya el servicio de recrear a seres queridos fallecidos e interactuar con ellos", 20 minutos, 20 bits, 3 de diciembre del 2023, (https://www.20minutos.es/tecnologia/empresas-inteligencia-artificial-ofrecen-servicio-recrear-seres-queridos-fallecidos-interactuar-con-ellos-5195903/) (recuperado el 1 de junio del 2024).

"Las obras creadas "exclusivamente" con Inteligencia Artificial no podrán ganar un Premio Nacional de Cultura" El Periódico de España, Cultura, noticias efe, 19 de febrero de 2024 (https://www.epe.es/es/cultura/20240219/obras-creadas-exclusivamente-inteligencia-artificial-codigo-buenas-practicas-ministerio-cultura-98374508) (recuperado el 13 de mayo del 2024).

"Marilyn conoce a James Bond en la película Deepfake del artista", El informe de Marilyn, 5 de marzo de 2024, "https://themarilynreport-com.translate.goog/2024/03/05/marilyn-meets-james-bond-in-artists-deepfake-movie/?_x_tr_sl=en&_x_tr_tl=es&_x_tr_hl=es&_x_tr_pto=sc (recuperado el 2 de agosto del 2024).

"El Comisionado de Hamburgo para la protección de Datos y la Libertad de Información pública en documento en el que se analiza los grandes Modelos de Lenguaje desde el punto de vista de la protección de datos, Boletín de julio del 2024", Lks, noticias, (https://www.lksnext.com/es/noticias_boletin/el-comisionado-de-hamburgo-para-la-proteccion-de-datos-y-la-libertad-de-informacion-publica-un-documento-en-el-que-analiza-los-grandes-modelos-de-lenguaje-desde-el-punto-de-vista-de-la-proteccion-de-d/) (recuperado el 2 de agosto del 2024).

Síganos en Linked

Visite nuestra web e infórmese de las novedades y actividades formativas que realizamos

www.rdu.es

2444-5762