EDITORIAL

Inteligencia artificial generativa

José Ramón Casar Corredera

Académico de Número y Presidente de la Sección de Ingeniería de la Real Academia de Doctores de España
joseramon.casar@upm.es

He aceptado con agrado la encomienda de la Junta de Gobierno para escribir el Editorial de este número de los Anales de la Real Academia de Doctores de España. Y he optado por hacerlo sobre un tema de actualidad en el ámbito de las ingenierías de la computación y la inteligencia artificial, que es la Inteligencia Artificial Generativa, es decir, ese tipo de inteligencia artificial capaz de generar contenidos, emulando lo que produciría un creador humano, y que ha supuesto una revolución viral a finales de 2022 y, sobre todo, durante 2023. Me propongo revisar algunos de los elementos principales de esta disciplina, como el aprendizaje automático no supervisado, los llamados modelos "fundacionales", la arquitectura de "transformador" y algunas otras, introducir algunos conceptos generales como la "explicabilidad" o la frugalidad, y considerar algunas aplicaciones, como la generación de texto en lenguaje natural o imágenes artísticas o de otro tipo; y también algunas de las posibilidades en medicina o educación.

Desafortunadamente, no tendré espacio para dedicar una mínima atención a determinados aspectos sociotécnicos o éticos ni al impacto profundo que este tipo de tecnología podrá implicar. Pero nuestra Academia se ha planteado celebrar en los próximos meses algunos encuentros sobre Inteligencia Artificial, que, con seguridad, cubrirán diversos dominios, y en los que tendremos ocasión, por tanto, de reflexionar sobre ello.

Para finalizar esta introducción, anticipo que este Editorial es un texto de revisión y de reflexión divulgativa, pero no un artículo científico-técnico ni un tutorial. No espere el lector encontrar en él todas las claves, ni siquiera un subconjunto relevante, ni todos los elementos constituyentes o tecnologías incumbidas. Como Editorial, sólo pretende poner en común en nuestra Academia, en su naturaleza multidisciplinar, el fenómeno de la emergencia inevitable de una tecnología que está impactando ya y que influirá progresiva y decisivamente en el desarrollo de nuestras áreas de conocimiento y profesiones, desde el Arte al Derecho y desde la Agricultura o la Industria a la Economía y la Medicina. Este

Editorial está escrito para un lector interesado, pero no especialmente versado en Inteligencia Artificial; que, por otro lado, tiene accesible fácilmente en internet una larguísima colección de recursos bibliográficos. Propondré, por mi parte, una breve lista de documentos seleccionados en relación con algunos (sólo algunos) de los aspectos discutidos, evitando, en lo posible, referirme a los excesivamente técnicos.

1.- LA INTELIGENCIA ARTIFICIAL GENERATIVA

Los métodos (no siendo radicalmente nuevos en su fundamento último, por cierto) y las aplicaciones de Inteligencia Artificial Generativa se han popularizado definitivamente en 2023. Etiquetamos con ese nombre al conjunto de métodos y aplicaciones capaces de generar contenidos (texto, imágenes, software o cualquier otra cosa) con características indistinguibles de las que produciría un ser humano. Para ello, esencialmente, las aplicaciones aprenden las características propias de los contenidos para las que han sido concebidas, a partir de una colección considerable de ejemplos reales, preferentemente de manera no supervisada, y terminan por ser capaces de producir nuevos contenidos con esas propiedades, con las instrucciones de generación que les pueda dar un usuario humano (instrucciones típicamente construidas en lenguaje natural o *prompts*). La muy reciente irrupción popular de ChatGPT o Bard (chat conversacional competidor, que se lanzó en marzo de 2023 por Google) han supuesto una revolución en la generación automática de contenidos. También el *chatbot* (o agente conversacional) del famoso buscador de Windows, Bing, incorpora ahora tecnología ChatGPT. Son aplicaciones, todas ellas, capaces de responder a una pregunta o generar un documento "consistente" como respuesta a cualquier demanda correctamente expresada en el dominio en el que se haya entrenado. Pero las aplicaciones de Inteligencia Artificial Generativa no se limitan a generar texto (o voz). Las de Visión y Arte Generativos, como DALL-E o MidJourney son capaces de generar imágenes originales con estilos aprendidos. Las posibilidades que ofrecen estas aplicaciones para acelerar la producción de contenidos valiosos son potencialmente inmensas, bajo la tutela y valoración de un humano experimentado (o sin ellas). Y las capacidades reales de la Inteligencia Artificial Generativa trascienden a los objetivos de generar texto informativo, educativo o conversacional o imágenes artísticas o publicitarias. Pueden producir también música "creativa" (como Jukebox de OpenAI o Music LM de Google), o código de software (como GitHubCopilot); o más allá, fórmulas de medicamentos, diseños de escenarios, argumentos legales o estrategias industriales (para la industria de los robots, los vehículos autónomos o la agricultura, por ejemplo). Mencionaremos, brevemente, algunas de estas áreas más abajo.

La Inteligencia Artificial Generativa, cuyos fundamentos técnicos, repito, no son nuevos, supone el paso de una visión prevalente (y aún válida) del razonamiento automático para percibir, clasificar o predecir a la de producir una imagen, un texto o, en general, un contenido. Esa generación no la hace desde la inteligencia innata, que obviamente no tiene, sino, como decía, después de aprender las características del dominio en el que se desee que opere, del modo en el que aprende un experto humano. El proceso implica, generalmente, un entrenamiento masivo, del que luego hablaremos brevemente y un ajuste sintonizado al dominio final de la aplicación (*fine-tuning*).

2.- LA INTELIGENCIA ARTIFICIAL GENERATIVA EN SU CONTEXTO HISTÓRICO Y SU EXPLOSIÓN RECIENTE

Se suele referir, como es sabido, el origen de la Inteligencia Artificial a los años 50 del siglo XX y concederle el crédito principal (con lo que estoy de acuerdo) al gran Alan Turing, a cuyo famoso test me referiré más tarde, al hilo de las llamadas Redes Neuronales Adversarias, y a la famosa conferencia de Dartmouth (*Dartmouth Summer Research Project on Artificial Intelligence)*, en la que participaron ni más ni menos que John McCarthy, Marvin Minsky, Claude Shannon, Allen Newely y Herbert Simon, entre otros. En esos setenta años (pongamos setenta por simplificar), se han producido notables avances en el área, se ha generado un corpus doctrinal y teórico y se han propuesto definiciones complementarias para los elementos centrales. Me quedo, para este Editorial, con dos definiciones de Inteligencia Artificial, como podría quedarme con otras o hacer referencia a otros conceptos.

Una podría ser esta de la Agencia Europea de Defensa: "Inteligencia Artificial es la capacidad de los algoritmos de seleccionar opciones óptimas o subóptimas, de entre un amplio espacio de posibilidades, para cumplir los objetivos pretendidos, aplicando diferentes estrategias, incluyendo la adaptación a las condiciones dinámicas del contexto y aprendiendo de la propia experiencia y de los datos suministrados o autogenerados".

Otra podría ser esta, de uno de los Centros de Excelencia del Gobierno de Estados Unidos, para definir un sistema inteligente: "Sistema artificial que realiza tareas, en circunstancias variables e impredecibles, sin supervisión humana relevante o que puede aprender de la experiencia y mejorar sus prestaciones cuando accede a determinados conjuntos de datos".

Impulsada por las aplicaciones en robótica, en agentes conversacionales, en el reconocimiento automático de imágenes y escenas y en varios e importantes ámbitos de la toma de decisiones, el área fue evolucionando, a pesar de los llamados "inviernos" de la Inteligencia Artificial. Efectivamente, entre mediados de los 60 y los últimos 90 del siglo XX, se produjo una recesión intelectual y científica en el ámbito, para resurgir desde entonces,

quizás primero con Deep Blue (una máquina de IBM, capaz de ganar al ajedrez a los mejores jugadores), con Watson, años más tarde (un ordenador, también de IBM, capaz de responder preguntas de toda índole) y, ya más recientemente, con la introducción de asistentes personales, como Siri o Alexa.

En el camino se fueron proponiendo una larga colección de aproximaciones y aplicaciones. Como no puedo referirme a todas, mencionaré sólo las primeras (y vigentes) herramientas simbólicas, como las ontologías, las redes semánticas o los sistemas expertos, que tratan de codificar el conocimiento humano en forma de reglas, y a la variante de los CBR (*Case Based Reasoning*, razonamiento basado en casos), que amplía el espacio de razonamiento de los sistemas expertos buscando casos similares, en su base de conocimiento, y aprendiendo de los resultados obtenidos, idealmente de forma continua. Estos sistemas, ciertamente, pueden aprender de la experiencia, pero no los calificaríamos (siempre) de sistemas de aprendizaje, con la perspectiva actual. En el ámbito de las herramientas no simbólicas (o menos simbólicas), necesariamente tenemos que mencionar el área de los multiagentes, que son sistemas constituidos por unidades software capaces de representar una unidad perceptiva-cognitiva de complejidad variable y de adaptar su comportamiento en relación con otros agentes para componer un sistema de razonamiento y de actuación (en ocasiones) global.

Y, finalmente, tenemos el aprendizaje automático o aprendizaje-máquina, que ha sido el desencadenante de las aplicaciones modernas de la Inteligencia Artificial y de la Inteligencia Artificial Generativa. El concepto se refiere a la asociación arquitecturas-algoritmos capaces de aprender las características de los datos de un dominio y, por tanto, de distinguir, de predecir y, en su caso, de generar de acuerdo con lo aprendido. Actualmente, la mayoría están basados en redes neuronales (y en particular en técnicas de aprendizaje profundo), pero provienen de algoritmos conocidos desde los 70 del siglo pasado, como el algoritmo de retropropagación o propagación hacia atrás (*backpropagation*), concebido para el llamado perceptrón, y también posiblemente de los trabajos de B. Widrow. Es difícil e inútil intentar dar el crédito especial a alguien. La invención es colectiva. Pero B. Widrow, con T. Hoff, propuso los algoritmos de ajuste de mínimos cuadrados para aprender los parámetros (de sistemas lineales) a partir de datos secuenciales; y G. Cybenko demostró teóricamente que una red neuronal con un nivel oculto puede aproximar cualquier función/región de decisión. Esas contribuciones no fueron menores, aunque hoy parezcan alejadas del actual estado del arte de la Inteligencia Artificial Generativa, construida sobre arquitecturas modernas de aprendizaje automático y de aprendizaje profundo, a las que dedicaré la próxima sección.

Efectivamente, la Inteligencia Artificial Generativa no sería posible sin determinados nuevos avances (junto con otros, como la capacidad de acceder a grandes volúmenes de datos estructurados y no estructurados y el desarrollo de las capacidades de cómputo

disponibles). En este punto, tenemos que recordar que se han venido generando datos, en varias disciplinas, desde hace mucho tiempo, con una diversidad de modelos, como los Modelos de Markov Ocultos (*Hidden Markov Models*) o los Modelos de Mezcla Gaussianos (*Gaussian Mixture Models*), entre otros. El cambio significativo para la Inteligencia Artificial Generativa se produce cuando se conciben (se generalizan) formas de algoritmos con capacidad de aprender de los ejemplos no supervisadamente.

El concepto arquitectural clave es el de red neuronal, que no es otra cosa que una red (un algoritmo) compuesta de neuronas artificiales, organizadas típicamente en niveles o capas (*layers*), desde uno de entrada hasta otro de salida, pasando por niveles intermedios (que puede ser uno o decenas). Cada neurona propaga su salida a otras a través de su "sinapsis", en donde se pondera por un peso y se transforma no linealmente por la llamada función de activación (por ejemplo, un sigmoide), con el objetivo de permitir mayor o menor grado de activación en la neurona receptora. Los pesos de cada una de las neuronas se ajustan con el objetivo de minimizar una función de coste global que representa el objetivo del problema, tradicionalmente usando una variante del algoritmo de gradiente, aunque no necesariamente con una propagación tradicional hacia atrás. El modelo se ha presentado tradicionalmente como uno que replica algunas estructuras básicas del cerebro humano, pero más allá de hasta qué punto eso es así o no, o si la analogía es interesante, su valor demostrado reside en su capacidad de generalizar y de aprender, de resolver nuevos problemas, abstrayendo a partir de los ejemplos de los que aprende.

3. CONCEPTOS TECNOLÓGICOS

El principal concepto que soporta la Inteligencia Artificial Generativa actual es el aprendizaje automático no supervisado sobre redes neuronales profundas, aunque también pueda utilizar el aprendizaje supervisado, sobre todo en la etapa de ajuste fino en un dominio concreto de aplicación. Con el término de aprendizaje supervisado nos referimos a aquel en el que la arquitectura aprende los valores correctos de los parámetros del modelo con la guía (humana o automática) de saber cuál es la respuesta correcta (sea en un problema de clasificación, de predicción, o en cualquier otro). Esto implica tener clasificados previa y correctamente los datos y tener "anotadas" previamente las respuestas. El aprendizaje no supervisado, por el contrario, comprende un conjunto de técnicas que permiten a la arquitectura aprender, sin referencia expresa previa de qué es correcto y qué no, es decir, directamente de los datos disponibles, sin un supervisor expreso. Evito en este Editorial hacer referencia al importante concepto de aprendizaje por refuerzo o a los denominados modelos de difusión, en los que está basado DALL-E, por ejemplo. Sólo el avance en los métodos de aprendizaje no supervisado, junto, como mencionaba más arriba,

con el desarrollo de la capacidad de cómputo y la creciente disponibilidad de datos masivos sobre los que aprender, ha permitido disponer de las actuales herramientas de Inteligencia Artificial Generativa, basadas prominentemente en los llamados Modelos Fundacionales.

Estos son, en esencia, arquitecturas de redes neuronales profundas pre-entrenadas exhaustivamente en algún dominio (como BERT y CLIP) con aprendizaje no supervisado y que representan, una vez entrenadas, un modelo multipropósito, que cada cual idealmente ajustará para su tarea y objetivos concretos. Los modelos actuales han sido desarrollados principalmente en el ámbito del llamado Procesado de Lenguaje Natural (es decir en tareas de predicción, interpretación y generación de lenguaje).

En buena medida, la Inteligencia Artificial Generativa está basada en estos supuestos y conceptos. Las arquitecturas son variadas y han evolucionado según el ámbito. Por ejemplo, en el área del Procesado de Lenguaje Natural, han progresado decisivamente desde las primitivas N-gram, pasando por las RNN (Redes Neuronales Recurrentes), a las arquitecturas actuales, para tener en cuenta la dependencia temporal en un texto, su carácter secuencial.

Sin embargo, en el ámbito del Reconocimiento (y la Generación) de Imágenes, las arquitecturas han evolucionado a partir de las CNN (Redes Neuronales Convolucionales), que se concibieron para procesar datos en forma de matriz o rejilla, principalmente para detectar patrones o formas en cualquier lugar de la imagen y, por tanto, especialmente útiles para clasificar, detectar o reconocer objetos en imágenes o en vídeos.

Pero posiblemente las redes de actualidad en relación al aprendizaje profundo para la Inteligencia Artificial Generativa sean las GAN (Generative Adversarial Networks, Redes Generativas Adversarias) y los VAE (Variational Autoencoders, Autocodificadores Variacionales), a las que quiero dedicar unos párrafos; especialmente a la primera, por su relación con la propuesta de Turing y su concepto adelantado de inteligencia artificial.

Las GAN son redes neuronales no supervisadas, pero basadas en un auto-entrenamiento supervisado, especialmente indicadas para la Inteligencia Artificial Generativa. Arquitecturalmente constan de un generador (una red neuronal que genera instancias a partir de lo que va aprendiendo en el dominio) y un discriminador, que trata de valorar, sin otro conocimiento, si la instancia que se le presenta es un ejemplar real extraído del mundo real o es un ejemplar simulado producido por el generador. Progresivamente, el generador va aprendiendo a generar ejemplos más realistas, más parecidos a los ejemplos reales, y el discriminador a distinguir mejor los casos verdaderos de los simulados, en un escenario de aprendizaje mutuo competitivo, en una suerte de lo que en Teoría de Juegos se denomina juego de dos jugadores de suma cero. Lo que uno gana, lo pierde el otro; o lo que uno acierta,

el otro lo falla. El resultado es que al final, en la situación de equilibrio, el generador aprende a generar datos con las características de los datos reales, o al menos con unas propiedades indistinguibles de los datos reales. Este es el fundamento de las GAN. Luego, el generador entrenado creará contenidos con las propiedades de los datos de partida, y, por tanto, será el agente básico de la Inteligencia Artificial Generativa.

No le hubiera dedicado tantas líneas en un artículo editorial como este a las GAN si no fuera porque, a mi juicio, representan, materializan y codifican la propuesta que hiciera Turing en 1950 sobre cómo evaluar la capacidad de una máquina para parecer (al menos) inteligente. La prueba propuesta consistía en presentar a un evaluador humano respuestas generadas por humanos y por máquinas. El evaluador debería determinar si la respuesta era de un humano o de una máquina. Alcanzado un nivel de "confusión" del evaluador, se podría inferir que la máquina "parecía" inteligente. El principio de operación de las GAN no es exactamente el mismo, pero comparte un principio filosófico común, a mi juicio.

La otra arquitectura relevante es la de los autocodificadores variacionales, que básicamente constan de un codificador capaz de representar las características principales del dominio de generación de contenidos, y un decodificador capaz de reconstruir instancias a partir de esa representación. El concepto subyacente, simplificando, es el de entender los datos, reducir su dimensión, y reconstruir nuevas formas a partir de esa representación.

Pero el salto actual se da con la adopción de los llamados transformadores y el mecanismo de "atención" propuesto por Varwani en 2017 (aunque existían antecedentes), que pondera la relevancia de cada palabra en la semántica y sentido general del texto, y por tanto, es capaz de capturar y generar contenido con sentido contextual. Las arquitecturas de transformador (que permiten paralelizar la "atención" y el cómputo) están compuestas también por codificadores y/o decodificadores, construidos con redes neuronales, como otras arquitecturas (con ambos o sólo de un tipo, más recientemente); los primeros encargados de mapear una secuencia de entrada en una cadena de símbolos y los segundos encargados de generar la secuencia de salida. Entre otros, el popular GPT-4 (*Generative Pre-Trained Transformer*) de OpenAI y BERT de Google están basados en estos conceptos.

Para terminar esta sección quiero hacer referencia breve a tres conceptos de interés general para la Inteligencia Artificial, y de interés también, algunos más y otros menos, para el campo de la Inteligencia Artificial Generativa actual, que ocupa a este Editorial. Son los de Explicabilidad, Interpretabilidad y Frugalidad. Haré una mención también a los de Transferencia de Conocimiento y Multimodalidad.

Los términos relacionados de Explicabilidad e Interpretabilidad hacen referencia a la necesidad, en algunas áreas al menos, de entender las decisiones propuestas o tomadas por

los algoritmos de Inteligencia Artificial. Cuando las soluciones propuestas están basadas en reglas codificadas es probablemente más fácil trazar el razonamiento e interpretar la solución. Pero con los modelos de aprendizaje profundo basado en redes neuronales de varias etapas resulta por lo general complejo, si no imposible. Cuando no se requiere conocer la razón, sino que sólo se desea precisión suficiente en una determinada predicción numérica o clasificación, por ejemplo, o disponer de contenidos creativos aceptables sobre las que un humano pueda juzgar positivamente sin más, estos conceptos pueden tener menos trascendencia (salvo si se desea adoptar una perspectiva epistemológica). Ese puede ser el caso de una parte de las aplicaciones de Inteligencia Artificial Generativa, pero no en áreas personal o socialmente sensibles como pueden ser el diagnóstico clínico u otras misiones críticas, como algunas de economía y de defensa, por poner dos ejemplos.

Los términos de Explicabilidad e Interpretabilidad suelen usarse indistintamente, aunque el término Interpretabilidad se refiere a la propiedad de entender la relación entre los datos procesados y la solución propuesta, en una perspectiva causa-efecto; y la Explicabilidad a la capacidad de entender el razonamiento efectuado por la máquina para llegar a las conclusiones presentadas o a las acciones propuestas.

La llamada Frugalidad tiene que ver con la propiedad de los algoritmos de aprendizaje de necesitar pocos recursos para mostrar unas prestaciones aceptables. Los recursos pueden ser datos, energía u otros (computacionales o de almacenamiento, por ejemplo); o, menos comúnmente, tiempo de entrenamiento, anotación o preprocesado supervisado. El concepto demandaría un artículo o editorial en sí mismo, que revisara las diferentes perspectivas posibles. Por ejemplo, la frugalidad en datos se refiere a la capacidad de generar resultados de calidad (contenidos, predicciones, clasificaciones, etc.) con pocos datos de los que aprender. Es un objetivo deseable siempre, pero especialmente relevante, como se comprende, cuando el dominio de aplicación ofrece pocos datos, por su naturaleza; pero también cuando los datos, uno a uno, requieren la atención específica y dedicada de un humano experimentado. La frugalidad en energía o en otros recursos se refiere a las necesidades de consumo de los modelos, tanto en la fase de entrenamiento en grandes servidores propios o en la nube, como en la fase de ejecución (esta segunda fase además crecientemente importante a medida que los algoritmos vayan implantándose en dispositivos personales portátiles).

Son centrales también los conceptos de Aprendizaje por Transferencia (*Transfer Learning*) y el de Multimodalidad. El primero se refiere a la posibilidad de utilizar, al menos como punto de partida, un modelo entrenado en una tarea para realizar otra, lo que es especialmente valioso cuando se dispone de pocos datos o pocos recursos para entrenar el modelo de interés propio. El segundo hace referencia a los modelos en los que se aprenden

y se usan tipos de datos de diferente representación, de diferente modalidad (series, imágenes, textos, etc.). Haré referencia breve a un caso más adelante en este Editorial.

4. ALGUNOS DOMINIOS DE APLICACIÓN DE LA INTELIGENCIA ARTIFICIAL GENERATIVA

Es sabido que la Inteligencia Artificial como tal tiene aplicación en muy diversos dominios: desde la codificación del conocimiento, la robótica avanzada, el reconocimiento e identificación de imágenes y actividades, a la predicción econométrica, la detección de anomalías o el mantenimiento predictivo; muchas veces, en combinación con métodos numéricos y lógicos clásicos.

Pero en estos meses, es la Inteligencia Artificial Generativa la que está despuntando mediática y prácticamente. Recuerdo al lector, innecesariamente, que con ese término nos referimos a los algoritmos específicos de aprendizaje no-supervisado y en el acceso a muchos datos, que son capaces de aprender y de generar salidas coherentes, consistentes y compatibles con lo que podría producir un ser humano entrenado (e imperfecto); al menos consistentes (o hasta cierto punto indistinguibles) desde la perspectiva propuesta por Turing a la que me refería anteriormente.

En esta sección, me propongo revisar brevemente (apenas unos párrafos por cada una de ellas) algunas de las áreas relevantes de aplicación de estos métodos: He elegido las aplicaciones de lenguaje y *chatbots* conversacionales, la generación de imágenes y el arte visual generativo en general (estas aplicaciones, inevitablemente, por el papel jugado en el desarrollo y la popularidad creciente de estos métodos); y después, entre las muchas posibilidades, he elegido escribir sobre algunas en Medicina y Biomedicina y en Educación. He dejado al margen las aplicaciones en defensa, en robótica, en producción de software y las aplicaciones de ocio y entretenimiento, de generación de publicidad personalizada, de coche autónomo, de arquitectura, urbanismo y diseño de espacios y otras varias.

4.a) El desarrollo de aplicaciones y sistemas conversacionales orientadas al lenguaje (*chatbots*) fueron el principal motivo que impulsó el avance de la Inteligencia Artificial Generativa en los últimos años. Son las que denominamos, como escribía más arriba, aplicaciones en el área de Procesado de Lenguaje Natural (comprensión, interpretación, generación, traducción).

En la semana en la que estoy escribiendo este Editorial, Meta (antes Facebook) ha anunciado el lanzamiento de SeamlessM4T, un modelo fundacional, capaz de trabajar con cien lenguajes distintos y de traducir de texto a voz o a la inversa y de voz a voz o de texto a texto.

También ha sacado LlaMa 2, otro modelo generativo de lenguaje, en cooperación con Microsoft y utilizando la tecnología de OpenAI. Todos los grandes actores tecnológicos han decidido apostar por esta tecnología, por proponer este modo de interacción con lenguaje natural y por este tipo de generación de contenidos. Habrá que monitorizar su recorrido, su futuro real y valorar su aceptación social.

4.b) Arte generativo:

DALL-E (Open AI) y Midjourney son dos aplicaciones, como otras varias, capaces de generar imágenes creativamente a partir de descripciones textuales del usuario, por ejemplo, imágenes artísticas y en estilos definidos con una determinada estética, a partir de imágenes aprendidas. Las GAN (y los VAE) son un instrumento para enseñar, quizás competitivamente, cómo generar imágenes en un determinado estilo, aunque el arte generativo en general, el arte algorítmico (en cierta manera serendípico) no está necesariamente basado en el aprendizaje automático; tradicionalmente ha venido incluyendo métodos de producción basada en reglas o en otros métodos generativos basados en fractales o en otras técnicas de tipo evolutivo y aleatorio. Las imágenes o representaciones tridimensionales de objetos que pueden generar las aplicaciones de Inteligencia Artificial generativa no tienen que ser aleatorias o abstractas; pueden ser figurativas y de expresividad realista.

Podríamos también mencionar aquí las aplicaciones para producir música a partir de patrones seleccionados aprendidos, vídeos inmersivos o escenarios y espacios virtuales.

4.c) Medicina y Biomedicina:

Es bien conocido el papel que está empezando a jugar la Inteligencia Artificial en la práctica médica, tanto en las fases relacionadas con el diagnóstico como en las de tratamiento. Ciertamente, muchas de las técnicas están todavía en fase experimental o de validación técnica y clínica, pero abarcan desde el apoyo a la prevención, la detección de anomalías, el uso de la imagen médica o la reconstrucción de órganos en tres dimensiones hasta la misma gestión de la interacción, tanto de pacientes como de médicos y sanitarios. La mayoría de estos sistemas pueden realizar tareas de detección, reconocimiento, clasificación, recomendación o comunicación avanzada.

Pues bien, los avances en las arquitecturas y modelos de la Inteligencia Artificial Generativa a los que nos hemos venido refiriendo en este Editorial no harán sino mejorar estas aplicaciones, con su capacidad de aprendizaje, e incorporarán otras. Posiblemente tres relevantes (no las únicas) a corto plazo serán:

- la capacidad de interacción textual (o vocal o de otro tipo o combinada) con los sistemas, que servirá para mejorar la comprensión y generar posibles diagnósticos y tratamientos, eficiente y razonadamente.

- La gestión de la multimodalidad, algo que no es exclusivo de la Inteligencia Artificial Generativa, pero que mejorará, muy probablemente, con sus estrategias de aprendizaje masivo. La gestión de la multimodalidad, incluida la consideración del propio contexto de la enfermedad y del paciente (con mecanismos de multi-atención por ejemplo) es crítica, creemos, en la práctica médica, en la que no sólo importan la descripción cualitativa textual de los síntomas o el análisis de un tipo de imágenes o análisis o la historia clínica separadamente, sino el conjunto.

- En Biomedicina y Farmacia, la Inteligencia Artificial Generativa permitirá diseñar nuevos medicamentos a base de generar "razonadamente" sustancias o compuestos a partir del conocimiento previo de sus propiedades y efectos, de la descripción de la enfermedad objetivo y de la experiencia adquirida. Esto no soslayará la necesidad de realizar ensayos y verificaciones, pero aportará la capacidad de generar rápida y eficientemente mejores medicamentos y con menores efectos secundarios.

En este hilo, hay al menos otros tres temas de interés (en el ámbito médico; pero no sólo en él); son los de explicabilidad (al que nos hemos referido anteriormente), responsabilidad y privacidad, que se describen por su propia denominación. Si un programa basado en inteligencia artificial diagnostica una enfermedad, debería ser capaz de explicar o al menos visualizar las evidencias en las que se ha basado, lo que no siempre es posible con los modelos profundos actuales de Inteligencia Artificial Generativa. Del mismo modo, si sugiere un tratamiento, debería poder justificarlo o quizás suministrar algunos datos de probabilidad de acierto (histórica). Y si el tratamiento se adopta, pero no tiene éxito, se añade un tema, nada menor, de atribución de responsabilidad, como es sabido. Finalmente, el tema de privacidad y consentimiento para el uso de los datos (que no parece, en principio el más complicado o, en principio, más complicado que en la actualidad) puede adoptar nuevos matices, como por ejemplo que admitamos mayoritariamente que nuestros datos completos se utilicen para entrenar otro sistema, del que no sabemos ni probablemente sabremos y que no nos beneficiará demostrablemente.

4.d) Educación

La Educación es otro de los grandes ámbitos en los que la Inteligencia Artificial Generativa tiene una posible ruta de éxito, dadas sus capacidades para producir contenidos educativos y metodologías adaptadas a cada caso, a partir de la cantidad de recursos disponibles de los que aprender. Y así está siendo reconocido reciente y repetidamente.

Entre otras posibilidades, se señala la de construir itinerarios de aprendizaje a medida, en los que los contenidos y las experiencias se personalizan en función de los intereses, necesidades y ritmo de progreso del estudiante; o la de producir material de aprendizaje interactivo, que puede ser a la vez pedagógico, eficaz y atractivo.

La Inteligencia Artificial Generativa también ofrece la posibilidad de incorporar la multimodalidad en el proceso de aprendizaje; es decir, la oportunidad de usar lo textual, lo visual, lo conversacional, respondiendo a preguntas y recomendando actividades, a modo de un tutor virtual. La actual tecnología al uso (que evolucionará, indefectiblemente) permite aprender de la inmensidad de recursos disponibles en internet, adaptarlos a un dominio de aprendizaje concreto y a un nivel de aprendizaje de referencia (supervisadamente, por un experto docente).

A pesar de su novedad, varios autores se han ocupado ya de valorar y opinar sobre el uso de esta inteligencia, y en especial de ChatGPT en Educación, ponderando sus ventajas e inconvenientes. Ciertamente, entre sus ventajas se encuentra la posibilidad de generar contenidos eficientemente y de calidad. Y, entre sus debilidades, podríamos mencionar que, con la tecnología actual, las respuestas que puede presentar el sistema al usuario carecen de argumentos, de evidencias. Pueden ser ciertas, correctas, útiles, pero no se explican (con frecuencia).

Desde nuestro punto de vista, existen aún algunos problemas básicos, que se refieren al modo en que entrenamos la red, con qué datos y con qué guías. Porque los datos abiertos y universalmente accesibles ofrecen una gran oportunidad informacional, pero el aprendizaje autosupervisado en temas que pueden afectar al carácter del aprendizaje o producir sesgos ideológicos o de otro tipo probablemente requieran una reflexión compartida y el desarrollo de una estrategia de evaluación. Y finalmente no es secundario el asunto de la aproximación pedagógica; una cosa es disponer de la capacidad de generar contenidos de calidad y otra determinar el modo de presentarlos y el contexto particular de la experiencia. Esto, sin duda, se irá abordando y resolviendo, pero, a día de hoy, creo (es una opinión personal) que podemos entender una inteligencia artificial en educación basada en reglas, pero no sé aún si podemos aceptar una educación basada en el aprendizaje ciego por una red neuronal de contenidos no seleccionados exhaustiva y cuidadosamente.

Luego, se han apuntado, por expertos y pensadores, otros problemas en Educación (y otras áreas), como el posible plagio automático de material ajeno o la vulneración, inconsciente o no, de derechos de propiedad intelectual en el proceso de aprendizaje y generación de contenidos o la propagación de materiales deficientes. En cualquier caso, es nuestra opinión que queda mucho por investigar y proponer en este dominio, que es el opuesto al tradicional

de la Inteligencia Artificial, en el que la persona enseña a la máquina, para convertirse en uno en el que la máquina enseña a la persona.

5. CONCLUYENDO

No tenemos espacio para revisar más aplicaciones en este Editorial. Nos hemos limitado a las de Procesado de Lenguaje Natural, de Generación de Arte Visual, en la Medicina y en la Educación. Dejamos para otra ocasión las de Defensa, Arquitectura, Diseño o Robótica. Sin embargo, hay una aplicación horizontal que no queremos dejar de mencionar para ir concluyendo, que es la aplicación para Aumento de Datos (*Data Augmentation)*. El objeto de esa perspectiva de la Inteligencia Artificial Generativa es la generación de datos sintéticos realistas con los que reentrenar y mejorar su propia capacidad generativa. Es un área clave en aquellos dominios en los que la disponibilidad de datos o no es grande o no es variada o no es representativa de determinados casos relevantes. Eso puede suceder en muchos ámbitos. Por ejemplo, en Medicina, las bases de datos de imágenes médicas, en general, ni son completas ni contienen suficientes ejemplares representativos de todos los casos de interés. En el caso de la Defensa los escenarios reales disponibles para evaluar determinados procesos de toma de decisiones o para entrenar son ciertamente escasos. Y lo mismo sucede en Arquitectura o Diseño de Espacios o de Objetos y en otras áreas.

Esa capacidad de producir ejemplares es intrínseca al concepto de Inteligencia Artificial Generativa. Queda por asegurar que los que se generan son los ejemplares adecuados y no se propagan los sesgos, si los hubiera, de los modos de generación.

Aunque conscientemente nos hemos limitado a hablar de Inteligencia Artificial, para terminar este Editorial tengo que referirme a la imposibilidad de separar de esta los avances y contribuciones de otras tecnologías concurrentes, entre las que debo mencionar la llamada Internet de la Cosas, y su capacidad de adquirir y comunicar datos, las tecnologías de Nube y Borde (*edge*), los métodos de interacción avanzada y otros avances en computación y disciplinas relacionadas, entre los que están la neurotecnología o las ingenierías neuromórfica y cuántica.

Creo que en el medio plazo lo que podemos esperar y desear es que avance la inteligencia artificial cooperativa Persona-Máquina. Es decir, una inteligencia artificial que propone y decide en determinados niveles, a instancias de un humano, y una inteligencia humana que determina, interactúa, inquiere y pide explicaciones antes de aceptar una decisión o de tomarla con los argumentos disponibles. Se han dado muchos pasos en esa dirección en los últimos muchos años, tanto en los métodos como en la forma de interactuar con los sistemas. Pero hay que seguir progresando, especialmente en ámbitos críticos de toma de

decisiones. En la Inteligencia Artificial Generativa, para progresar en ese campo de cooperación Persona-Máquina, tenemos que acometer la mejora de las capacidades de interpretación, por parte de la máquina, del texto o del contenido generado por ella, porque con frecuencia no es capaz de expresar argumentos sino sólo de presentar estadísticas (en el mejor de los casos).

La Inteligencia Artificial no es una tecnología milagrosa, no es una magia, son algoritmos numéricos y conjuntos de datos de los que aprenden. Las respuestas de cualquier chat conversacional están basadas en lo que han aprendido. La clave, por tanto, son los datos de entrenamiento. Hay que elegirlos selectivamente, cosa que no sabemos si hacen los actuales modelos fundacionales. Otra clave es la "explicabilidad"; porque la explicación, la argumentación es la base de la confianza que podremos depositar en esos sistemas.

Quedan algunos caminos que andar y algunos destinos por descubrir. La Inteligencia Artificial Generativa tiene que aprender de su propia experiencia de manera natural, tanto en sus fases de interacción y creación como en su fase de evaluación. En cualquier caso, estoy seguro de que estamos en transición hacia otros paradigmas, que aún no me atrevo a anticipar.

6. ALGUNAS REFERENCIAS BIBLIOGRÁFICAS

- DAVID M. PATEL, Artificial Intelligence & Generative AI, Printed by Amazon, July 2023
- YIHAN CAO et al., A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT, J. ACM, Vol. 37, No. 4, August 2018
- MCKINSEY, What is generative AI? January 19, 2023 https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai
- KPMG, Inteligencia Artificial Generativa, https://www.tendencias.kpmg.es/2023/06/inteligencia-artificial-generativa-apocalipticos-vs-integrados/
- ASHISH VASWANI et al., Attention Is All You Need, 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, https://arxiv.org/pdf/1706.03762.pdf
- ALAN TURING, Computing Machinery and Intelligence, *Mind*, Volume LIX, Issue 236, October 1950, Pages 433–460, https://doi.org/10.1093/mind/LIX.236.433
- RISHI BOMMASANI et al., On the Opportunities and Risks of Foundation Models, https://www.researchgate.net/publication/353941945_On_the_Opportunities_and_Risks_of_Foundation_Models

- JASON BROWNLEE, A Gentle Introduction to Generative Adversarial Networks (GANs), June 17, 2019, https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/
- PHILIP GALANTER, Artificial Intelligence and Problems in Generative Art Theory, July 2019, Proceedings of EVA London 2019 (EVA 2019), pp 112-118
- PALADUGU, P.S., ONG, J., NELSON, N. *et al.* Generative Adversarial Networks in Medicine: Important Considerations for this Emerging Innovation in Artificial Intelligence. *Ann Biomed Eng* (2023). https://doi.org/10.1007/s10439-023-03304-z.
- BCG, Generative AI Will Transform Health Care Sooner Than You Think, June 22, 2023, https://www.bcg.com/publications/2023/how-generative-ai-is-transforming-health-care-sooner-than-expected
- GRANT COOPER, Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence, Published online: 22 March 2023, Journal of Science Education and Technology (2023) 32:444–452 https://doi.org/10.1007/s10956-023-10039-
- LINARDATOS P. et al., Explainable AI: A Review of Machine Learning Interpretability Methods. Entropy (Basel). 2020 Dec 25;23(1):18. doi: 10.3390/e23010018. PMID: 33375658; PMCID: PMC7824368
- LAURA SARTORI, ANDREAS THEODOROU, A sociotechnical perspective for the future of AI: narratives, inequalities, and human control, Ethics and Information Technology (2022) 24:4, https://doi.org/10.1007/s10676-022-09624-3