



CENTRO
INNOVACIÓN Y
EMPREDIMIENTO

ChatGPT E INTELIGENCIA ARTIFICIAL GENERATIVA: OPORTUNIDADES Y DESAFÍOS

JUNIO | 2023

En solo cinco días, un millón de usuarios acudieron a ChatGPT, el modelo de lenguaje de inteligencia artificial (IA) generativa de OpenAI que crea contenido original en respuesta a las solicitudes de los usuarios. Apple tardó más de dos meses en alcanzar el mismo nivel de adopción para su iPhone. Facebook tuvo que esperar diez meses y Netflix más de tres años para construir la misma base de usuarios.

Los casos de uso de IA generativa (por ejemplo, ChatGPT) ya están tomando vuelo a través de las industrias. El gigante de servicios financieros Morgan Stanley está probando la tecnología para ayudar a sus asesores financieros a aprovechar mejor los conocimientos de más de 100.000 informes de investigación de la empresa, mientras que el gobierno de Islandia se ha asociado con OpenAI en sus esfuerzos por preservar la lengua islandesa en peligro. Por su parte, Salesforce ha integrado la tecnología en su popular plataforma de gestión de relaciones con el cliente (CRM).

Este artículo de McKinsey (2023)¹, pretende proporcionar una comprensión del ecosistema que ha generado la IA generativa, su cadena de valor y las futuras oportunidades de inversión. El informe se basa en investigaciones primarias y secundarias, incluidas más de 30 entrevistas con fundadores de empresas, directores ejecutivos, científicos principales y líderes empresariales que trabajan para comercializar la tecnología; cientos de informes de mercado y artículos; y datos de investigación de McKinsey.

Una breve explicación de la IA generativa

Es útil tener un conocimiento básico de lo que es la IA generativa y cómo sus capacidades difieren de las tecnologías de IA "tradicionales" que las empresas utilizan para, por ejemplo, predecir la pérdida de clientes, pronosticar la demanda de productos, y hacer recomendaciones de productos. Una diferencia clave es su capacidad para crear nuevos contenidos. Este contenido puede ser entregado en múltiples modalidades, incluyendo texto (como artículos o respuestas a preguntas), imágenes que parecen fotos o pinturas, videos y representaciones en 3D (como escenas y paisajes para videojuegos).

Todo esto es posible mediante el entrenamiento de redes neuronales (un tipo de algoritmo de machine learning) en enormes volúmenes de datos y la aplicación de "mecanismos de atención", una técnica que ayuda a los modelos de IA a entender en qué centrarse. Con estos mecanismos, un sistema generativo de IA puede identificar patrones de palabras, relaciones y el contexto de la solicitud de un usuario.

¹ Este resumen se basa en el artículo "Exploring opportunities in the generative AI value chain", de Härlin, T., Sokolov, O. y Sukharevsky, A. T. MIT Sloan Management, 2023. Disponible en: https://www.mckinsey.com/capabilities/quantumblack/our-insights/exploring-opportunities-in-the-generative-ai-value-chain?stcr=FC95AF184E5549989A4964ADA05740C4&cid=other-eml-alt-mip-mck&hlkid=03d075186373440193b7fca153726600&hctky=12072984&hdpid=886dcb5f-5334-46d2-9be1-7cde6e362ecf#

La inteligencia artificial tradicional también podría usar redes neuronales y mecanismos de atención, pero estos modelos no están diseñados para crear nuevo contenido. Solo pueden describir, predecir o prescribir algo basado en contenido existente.

Cadena de valor de la IA generativa

A medida que el desarrollo y la implementación de sistemas de IA generativa se pone en marcha, una nueva cadena de valor está surgiendo para apoyar la formación y el uso de esta poderosa tecnología. Los fundamentos de los sistemas de IA generativa son considerablemente más complejos que la mayoría de los sistemas de IA tradicionales. En consecuencia, el tiempo, el costo y la experiencia asociados con la entrega de los sistemas dan lugar a importantes contratiempos para los nuevos participantes y pequeñas empresas en gran parte de la cadena de valor.

Muchas áreas seguirán siendo dominadas por gigantes tecnológicos en el futuro cercano. El mercado de aplicaciones de IA generativa es la sección de la cadena de valor que se espera que se expanda más rápidamente y ofrezca importantes oportunidades de creación de valor tanto a las empresas tecnológicas existentes como a los nuevos participantes en el mercado. Las empresas que utilizan datos especializados o patentados para ajustar las aplicaciones pueden lograr una ventaja competitiva significativa sobre las que no lo hacen.

Existen 6 categorías en la cadena de valor de IA generativa que serán importantes en el futuro: hardware de computadora, plataformas en la nube, modelos básicos, centros de modelos y operaciones de aprendizaje automático (MLOps), aplicaciones y modelos de base solo para servicios son una nueva adición. Estas categorías se describen a continuación:

- *Hardwares de computadora*

Los sistemas de IA generativa necesitan conocimiento, y mucho de él, para crear contenido. El GPT-3 de OpenAI, el modelo de IA generativa que sustenta el ChatGPT, por ejemplo, fue entrenado en unos 45 terabytes de datos de texto. No es algo que el hardware tradicional pueda manejar. Estos tipos de cargas de trabajo requieren grandes grupos de unidades de procesamiento gráfico (GPU) o unidades de procesamiento tensor (TPU) con chips "aceleradores" especializados capaces de procesar todos esos datos a través de miles de millones de parámetros en paralelo.

Una vez completado el entrenamiento de este modelo de IA generativa fundacional, las empresas también pueden usar estos clusters para personalizar los modelos (un proceso llamado "tuning") y ejecutar estos modelos hambrientos de energía dentro de sus aplicaciones. Sin embargo, en comparación con el entrenamiento inicial, estos últimos pasos requieren mucho menos poder computacional.

- *Modelos básicos*

Estos grandes modelos de aprendizaje profundo están preparados para crear un tipo particular de contenido y

pueden adaptarse para soportar una amplia gama de tareas. Una vez que se desarrolla el modelo de base, cualquiera puede construir una aplicación encima de ella para aprovechar sus capacidades de creación de contenido. Considere los modelos de fundación GPT-3 y GPT-4 de OpenAI que pueden producir texto de calidad humana. Suministran energía a docenas de aplicaciones, desde el tan comentado chatbot ChatGPT hasta los generadores de contenido de software como servicio (SaaS) Jasper y Copy.ai.

El desarrollo de modelos básicos requiere una amplia experiencia en varias áreas. Estos incluyen preparar los datos, seleccionar la arquitectura del modelo que puede crear la salida específica, entrenar el modelo, y luego ajustar el modelo para mejorar la salida (lo que implica etiquetar la calidad de la salida del modelo y alimentarlo de nuevo en el modelo para que pueda aprender).

- *Plataformas en la nube*

Las GPU y TPU son caras y escasas, lo que dificulta y no es rentable para la mayoría de las empresas adquirir y mantener esta plataforma de hardware vital en las instalaciones. Como resultado, gran parte del trabajo para construir, ajustar y ejecutar grandes modelos de IA se produce en la nube. Esto permite a las empresas acceder fácilmente a la energía computacional y gestionar su gasto según sea necesario. Como era de esperar, los principales proveedores de nube tienen las plataformas más completas para ejecutar cargas de trabajo de IA generativa y acceso preferencial al hardware y los chips.

Los competidores especializados en la nube podrían ganar cuota de mercado, pero no en un futuro próximo y no sin el apoyo de una gran empresa que busca reducir su dependencia de hiperescaladores.

- *Modelos hubs y MLOps*

Para crear aplicaciones basadas en modelos básicos, las empresas necesitan dos condiciones. La primera es un lugar para almacenar y acceder al modelo de base. En segundo lugar, es posible que necesiten herramientas, tecnologías y prácticas especializadas para adaptar un modelo básico e implementarlo en sus aplicaciones de usuario final. Esto incluye, por ejemplo, capacidades para incorporar y etiquetar datos de capacitación adicionales o crear las API que permiten a las aplicaciones interactuar con ellos.

Los centros de modelos proporcionan estos servicios. Para los modelos de código cerrado en los que el código fuente no se pone a disposición del público, el desarrollador del modelo de fundación suele servir como un hub de modelos. Ofrecerá acceso al modelo a través de una API a través de un acuerdo de licencia. A veces, el proveedor también ofrecerá capacidades MLOps para que el modelo se pueda ajustar e implementar en diferentes aplicaciones.

Hay muchas maneras en que los proveedores de aplicaciones pueden crear valor. Al menos en el corto plazo, vemos una categoría de aplicaciones que ofrecen el mayor potencial para la creación de valor. Y esperamos que las aplicaciones desarrolladas para ciertas industrias y funciones proporcionen más valor en los primeros días de la IA generativa.

- *Aplicaciones creadas a partir de modelos afinados*

En términos generales, encontramos que las aplicaciones de IA generativa caen en una de dos categorías. La primera representa los casos en los que las empresas utilizan modelos básicos en gran medida como es dentro de las aplicaciones que construyen, con algunas personalizaciones. Estos podrían incluir la creación de una interfaz de usuario a medida o la adición de orientación y un índice de búsqueda de documentos que ayudan a los modelos a entender mejor las indicaciones comunes del cliente para que puedan devolver una salida de alta calidad.

La segunda categoría representa la parte más atractiva de la cadena de valor: aplicaciones que aprovechan modelos básicos ajustados -aquellos que han recibido datos relevantes adicionales o han ajustado sus parámetros- para producir resultados para un caso de uso particular. Si bien los modelos de formación básica requieren cantidades masivas de datos, es extremadamente costoso y puede llevar meses, los modelos de base de ajuste requieren menos datos, cuestan menos y se pueden completar en días, poniéndolo al alcance de muchas empresas.

Las organizaciones también podrían aprovechar los datos de propiedad de las operaciones comerciales diarias. Un desarrollador de software que ha sintonizado un chatbot de IA generativo específicamente para los bancos, por ejemplo, puede asociarse con sus clientes para incorporar datos de chats de centros de llamadas, lo que les permite elevar continuamente la experiencia del cliente a medida que crece su base de usuarios.

En todos los casos, los desarrolladores de aplicaciones tendrán que vigilar los avances de la IA generativa. La tecnología se está moviendo a un ritmo rápido, y los gigantes tecnológicos continúan implementando nuevas versiones de modelos de bases con capacidades aún mayores. OpenAI, por ejemplo, informa que su recientemente introducido GPT-4 ofrece "conocimientos generales más amplios y habilidades para resolver problemas" para una mayor precisión. Los desarrolladores deben estar preparados para evaluar los costos y beneficios de aprovechar estos avances dentro de su aplicación.

Identificar la primera ola de impacto de aplicaciones por función e industria

Si bien es probable que la IA generativa afecte a la mayoría de las funciones empresariales a largo plazo, nuestra investigación sugiere que la tecnología de la información, el marketing y las ventas, el servicio al cliente y el desarrollo de productos están más maduros para la primera ola de aplicaciones.

- *Tecnología de la información*

La IA generativa puede ayudar a los equipos a escribir código y documentación. Ya, los programadores automatizados en el mercado han mejorado la productividad de los desarrolladores en más del 50%, ayudando a acelerar el desarrollo de software.

- *Marketing y ventas*

Los equipos pueden utilizar aplicaciones de IA generativa para crear contenido para el alcance del cliente. Dentro de dos años, se espera que el 30% de todos los mensajes de marketing salientes se desarrollen con la ayuda de sistemas de IA generativa.

- *Atención al cliente*

Los chatbots y asistentes virtuales personalizados de sonido natural pueden manejar las consultas de los clientes, recomendar una resolución rápida y guiar a los clientes a la información que necesitan. Empresas como Salesforce, Dialpad y Ada ya han anunciado ofertas en esta área.

- *Desarrollo de productos*

Las empresas pueden utilizar IA generativa para rápidamente prototipos de diseños de productos. Las empresas de ciencias de la vida, por ejemplo, ya han comenzado a explorar el uso de la IA generativa para ayudar a generar secuencias de aminoácidos y nucleótidos de ADN para acortar la fase de diseño de medicamentos de meses a semanas.

- *Servicios*

Al igual que con la IA en general, surgirán servicios de IA generativa dedicados para ayudar a las empresas a llenar vacíos de capacidad mientras compiten por desarrollar su experiencia y navegar por las oportunidades de negocio y las complejidades técnicas. Se espera que los proveedores de servicios de IA existentes desarrollen sus capacidades para servir al nuevo mercado.

Las empresas de nicho también pueden ingresar al mercado con conocimientos especializados para aplicar IA generativa dentro de una función específica (como cómo aplicar IA generativa a los flujos de trabajo de servicio al cliente), la industria (por ejemplo, guiar a las empresas farmacéuticas sobre el uso de la IA generativa para el descubrimiento de medicamentos), o capacidad (como cómo construir bucles de retroalimentación eficaces en diferentes contextos).

Si bien la tecnología de IA generativa y su ecosistema de apoyo todavía están evolucionando, ya está bastante claro que las aplicaciones ofrecen las oportunidades de creación de valor más significativas. Aquellos que pueden aprovechar los datos de nicho o, incluso mejor, de propiedad en los modelos de base de ajuste fino para sus aplicaciones pueden esperar lograr la mayor diferenciación y ventaja competitiva. La carrera ya ha comenzado, como lo demuestra el flujo constante de anuncios de proveedores de software, tanto existentes como nuevos participantes en el mercado, que traen nuevas soluciones al mercado.