



Facultad de Ciencias y Estudios Tecnológicos
Escuela de Estudios de Posgrado



EL JUEGO DE TRONOS EN LA ERA TECNOLÓGICA POR LA HEGEMONÍA EN LA INTELIGENCIA ARTIFICIAL

Dr. Roberto Morales Estrella

Profesor Investigador de la UAH

23 de septiembre de 2024



En 1996 se publicó por primera vez la novela de ciencia ficción, el Juego de Tronos, escrita por George Raymond Richard Martin (Newsletter, CCCB, 2012), la trama se centra en la lucha por el dominio de los siete reinos, destacándose las pasiones humanas, derivando en una sociedad caótica.

La comparación con lo que se está registrando en la competencia por el dominio del mercado de la Inteligencia Artificial Generativa (IAG), es similar en cuanto a la intensidad de la lucha por lograr la hegemonía de los mercados, ya que quien domine los mercados de la IA, impondrá las reglas del juego económico, financiero, tecnológico y sobre todo social, porque sino hay la regulación adecuada, a nivel de consenso mundial, predominarán los criterios del mercado libre, que es la rentabilidad por encima de lo social y lo humano

Fundada en 1993 Nvidia, logró notoriedad, con el ChatGPT-3, que lanzó el 30 de noviembre del 2022 (Osornio & Edwards, Bend, 2023), registrando resultados espectaculares en poco tiempo, en virtud de hacer visibles los modelos de lenguaje en extenso o lenguaje natural, fue el despegue de la Inteligencia Artificial Conversacional o IAG, convirtiendo estos modelos en una herramienta de gran aceptación, que posicionó a Nvidia como líder en el mercado de la IAG

Nvidia también ha logrado grandes avances con sus Unidades de Procesamiento Gráfico (GPU) las cuales son altamente reconocidas en los mercados, como excelentes para aplicaciones de IAG, además de contar con una arquitectura, que hace de sus GPU un hardware informático general, para aplicaciones de renderización de gráficos, con capacidades de aceleración de las cargas de trabajo, implementados con potentes modelos como H100, H200 y B200, fortaleciendo a la GPU con la plataforma denominada CUDA para apoyar a los desarrolladores, optimizando el aprovechamiento del hardware.

Una de las estrategias de Nvidia fue aprovechar a los estudiantes de las universidades, con las que estableció vinculación, para capacitarlos en la programación de CUDA según Matt Kimbal del Centro de Datos en Moor Insights & Strategy (Smith, 2024).

Dentro de los vectores clave de Nvidia, destaca la reestructuración de un billón de dólares de infraestructura de centros de datos, que se proyecta crecer a dos billones de dólares, para atender un mundo de computación acelerada, en lo referente a IAG, los otros vectores son las transiciones



generacionales en computación, la amplitud de la base de clientes, y las continuas limitaciones de suministros (Staff Boletín, 2024), sin lugar a dudas que los chips para IAG de Nvidia son tan escasos como demandados, por lo que ha diseñado un hoja de ruta para reavivar el negocio en China. Sus ingresos han pasado de 12 mmdd en 2019 a 60 mmdd en 2024, (ob.cit.)

Los competidores de Nvidia son empresas muy fuertes en I+D, como AMD con su modelo de GPU denominado Instinct-MI325K, competidor del H100 de Nvidia y su plataforma ROCm; Intel es el otro jugador con una plataforma de programación unificada de código abierto, llamada OneAPI con la cual contempla múltiples categorías de hardware; También entra a la competencia la empresa Cerebras participando en el mercado con sus chips en tamaño de obleas como el denominado Wafer Scale Engine, fabricado con silicio del tamaño de una oblea, pero con 1,2 billones de transistores de gran rendimiento y capacidad de memoria.

SambaNova fundada en 2017 ha creado su chip de entrenamiento SN40L con una arquitectura de flujo de datos reconfigurable, basada en mosaicos de memoria, además de su plataforma Samba Flow para analizar modelos de aprendizaje.

Cabe mencionar a Groq que opera convergiendo los recursos de memoria y computación para acelerar la velocidad con la que los modelos de lenguaje en extenso puede responder a las indicaciones.

La guerra entre los corporativos tecnológicos tienen como principales armas la I+D+i y las fuertes inversiones, como la Asociación Global de Inversión en Infraestructura de IA (GAIIIP) integrada por Microsoft, BlackRock, Global Infrastructure Partners (GIP) y MGX, que tienen como objetivo recaudar 100 mmdd para invertir en IA.

Lo preocupante de este panorama, es la ausencia del menor sentido social y de los derechos humanos, que los gobiernos deben establecer como prioridad, de lo contrario ¿adonde vamos?

Bibliografía

Newsletter, CCCB. (28 de julio de 2012). Centro de Cultura Contemporánea de Barcelona. Recuperado el septiembre de 2024, de CCCB: <https://www.cccb.org/es/multimedia/videos/george-r-r-martin/211363>

Osornio, A., & Edwards, Bend. (30 de noviembre de 2023). WIRED. Recuperado el septiembre de 2024, de es.wired.com: <https://es.wired.com/articulos/ano-uno-de-chatgpt-asi-ha-cambiado-el-mundo-desde-la-llegada-del-chatbot-de-openai>

Smith, M. S. (16 de septiembre de 2024). Los Rivaless se acercan para disputarles la corona a Nvidia. IEEE Spectrum.

Staff Boletín. (23 de febrero de 2024). Boletín de la Computación. Recuperado el septiembre de 2024, de boletin.mx: <https://boletin.mx/2024/02/23/nvidia-y-su-posicionamiento-en-la-actualidad-financiera-y-tecnologica/>