

Deepseek

Pasquel Johann

Tabla de Contenidos

GITHUB	1
1. OBJETIVOS	1
2. INTRODUCCIÓN	2
Distillation	2
3. DESARROLLO	3
Robo de datos a OpenAI	3
Costo de entrenamiento	3
EE.UU	3
GPUs	4
4. CONCLUSIONES	4
5. REFERENCIAS BIBLIOGRÁFICAS	4

GITHUB

https://github.com/Vladimirjon/MetodosNumericos_PasquelJohann/tree/main/Extra_MN/Deepseek

1. OBJETIVOS

- Investigar la relación entre el auge de DeepSeek y el contexto político global, especialmente su impacto en la estrategia tecnológica de Estados Unidos frente a China.

- Analizar en profundidad la técnica de distillation utilizada por DeepSeek y evaluar las acusaciones de robo de datos a OpenAI, explorando las implicaciones éticas, legales y de propiedad intelectual que se derivan de estas prácticas.

2. INTRODUCCIÓN

Distillation

En el desarrollo de modelos de inteligencia artificial, una de las técnicas más utilizadas para optimizar el rendimiento y reducir costos computacionales es la destilación de modelos (*distillation*). Este método permite transferir el conocimiento de un modelo de gran escala (*denominado teacher model*) a uno más pequeño y eficiente (*student model*), con el objetivo de mantener un rendimiento similar mientras se reduce la demanda de recursos computacionales.

Sin embargo, en el caso de la empresa **DeepSeek**, el uso de esta técnica ha generado **controversia debido a sospechas** de que su modelo podría haber sido entrenado a partir de datos generados por OpenAI, lo que plantea preocupaciones en torno a la ética y la legalidad de su implementación.

La destilación de modelos no es un proceso nuevo en la inteligencia artificial. Empresas como Google, Microsoft y Meta han empleado este enfoque para mejorar la eficiencia de sus modelos sin comprometer significativamente su capacidad de respuesta.

No obstante, cuando la destilación se realiza a partir de datos obtenidos de manera no autorizada, puede derivar en problemas de propiedad intelectual y derechos sobre los datos. En este contexto, **OpenAI y Microsoft han iniciado investigaciones para determinar si DeepSeek ha replicado información de sus modelos sin permiso**, aprovechando indirectamente los recursos y la inversión de estas compañías en el desarrollo de inteligencia artificial avanzada.

Algunas de las técnicas más utilizadas en este proceso incluyen:

- **Destilación basada en logits:** Se entrena el modelo pequeño para replicar la distribución de probabilidades de salida del modelo original.
- **Destilación a nivel de características:** Se transfieren representaciones intermedias de los datos en lugar de solo las salidas finales.
- **Destilación de secuencia a secuencia:** Se optimiza el aprendizaje en modelos de lenguaje y traducción, asegurando que el modelo pequeño imite la generación de texto del modelo grande.
- **Destilación de refuerzo:** Se utiliza en sistemas donde la IA aprende a través de interacciones con su entorno, aplicando técnicas de aprendizaje por refuerzo.

3. DESARROLLO

Robo de datos a OpenAI

DeepSeek ha sido objeto de controversia debido a las denuncias que sostienen que la empresa se habría beneficiado de técnicas de distillation utilizando datos obtenidos de manera no autorizada a través de la API de OpenAI.

Dichas alegaciones imponen una disertación crítica sobre la ética en el manejo y la transferencia de información, argumentando que este supuesto robo de datos podría violar principios fundamentales de propiedad intelectual y confidencialidad, elemento esencial para la confianza en el desarrollo tecnológico en inteligencia artificial.

El método que podría haber facilitado esta supuesta apropiación es la destilación de modelos, una técnica que permite entrenar modelos más pequeños utilizando las respuestas de modelos más avanzados.

A través de este proceso, un modelo puede aprender patrones, estructuras de lenguaje y comportamientos sin necesidad de acceder directamente a los pesos o parámetros del modelo original.

Costo de entrenamiento

El costo de entrenamiento de DeepSeek se ha reducido gracias a estrategias de optimización, que incluyen el uso eficiente de hardware y técnicas de aprendizaje avanzado.

Durante el preentrenamiento, el modelo procesó cada billón de tokens utilizando 180.000 horas de GPU H800, completando este proceso en menos de dos meses con 2.788.000 horas de GPU en total.

Suponiendo un costo de alquiler de 2 dólares por hora de GPU, el gasto final se estima en aproximadamente 5,576 millones de dólares.

EE.UU

El desarrollo de modelos de inteligencia artificial en China, particularmente los de DeepSeek, ha generado preocupación en el ámbito político y tecnológico de Estados Unidos.

La creciente capacidad de las empresas chinas en el sector de la IA se ha convertido en un tema estratégico para la administración estadounidense, influyendo en debates sobre seguridad nacional, competencia tecnológica y control de exportaciones de hardware avanzado.

En un contexto de tensión tecnológica entre **Estados Unidos y China**, la aparición de DeepSeek no solo representa un desafío comercial, sino también un factor que podría influir

en futuras decisiones de política internacional, incluyendo restricciones comerciales, sanciones tecnológicas y acuerdos de cooperación en el ámbito de la inteligencia artificial.

GPUs

Según informes de SemiAnalysis, DeepSeek cuenta con una infraestructura que incluye alrededor de **50.000 GPUs** de la arquitectura Hopper de NVIDIA, distribuidas entre los modelos H100, H800 y H20.

Las H100 son actualmente las más avanzadas en términos de capacidad de cómputo para inteligencia artificial, pero debido a las restricciones de exportación impuestas por el gobierno estadounidense, China ha recurrido al uso de modelos alternativos como la H800 y H20, diseñadas para cumplir con los límites establecidos por las regulaciones comerciales.

El acceso de DeepSeek a estas GPUs ha sido un factor clave en su capacidad de entrenamiento y ha generado debate sobre la efectividad de las sanciones tecnológicas impuestas por Estados Unidos.

Además de las GPUs de NVIDIA, algunos informes sugieren que **DeepSeek está explorando alternativas en chips de fabricación china**, como los desarrollados por Huawei y Biren Technology, lo que podría reducir su dependencia de proveedores estadounidenses en el futuro. Esta tendencia refleja una estrategia a largo plazo para fortalecer la industria tecnológica china y minimizar los efectos de las sanciones impuestas por Estados Unidos.

4. CONCLUSIONES

- El costo de entrenamiento de DeepSeek, aunque declarado como significativamente menor, ha sido objeto de debate debido a la magnitud de su infraestructura de GPUs, lo que pone en duda la transparencia en la inversión tecnológica y el acceso a hardware especializado.
- La posible apropiación de datos de OpenAI por parte de DeepSeek mediante técnicas de destilación plantea un desafío en la protección de la propiedad intelectual en inteligencia artificial, evidenciando la necesidad de regulaciones claras sobre el uso de datos generados por modelos preexistentes.

5. REFERENCIAS BIBLIOGRÁFICAS

<https://tech.slashdot.org/story/25/01/28/1315215/deepseek-has-spent-over-500-million-on-nvidia-chips-despite-low-cost-ai-claims-semianalysis-says>

<https://semianalysis.com/2025/01/31/deepseek-debates/>

<https://www.tomshardware.com/tech-industry/artificial-intelligence/deepseek-might-not-be-as-disruptive-as-claimed-firm-reportedly-has-50-000-nvidia-gpus-and-spent-usd1-6-billion-on-buildouts>