

2025-02-28

Los datos, fuente de inteligencia

Jonathan Gómez

#IA

#IAGenerativa

Resumen

La inteligencia artificial depende fundamentalmente de los datos, que son la base de su aprendizaje y rendimiento. El aprendizaje profundo y los LLM procesan grandes cantidades para generar contenido diverso. Conceptos clave como IA, aprendizaje máquina, aprendizaje profundo y modelos generativos son esenciales para comprender este campo versátil. Si bien la IA ofrece grandes aplicaciones, el tratamiento ético y legal de los datos, especialmente los personales, es crucial para equilibrar innovación y privacidad.

Introducción

La gran cantidad de información producida por el humano es enorme. Según el sitio web [statista.com](https://www.statista.com), se prevé que la creación de datos a nivel mundial para el 2028 crezca a más de 394 zettabytes. Es importante mencionarlo, se puede pensar que los resultados obtenidos en inteligencia artificial únicamente dependen de aspectos meramente computacionales o algorítmicos, sin embargo sin importar la complejidad de la arquitectura que se use para entrenar una red neuronal los datos son parte fundamental para que rinda frutos tal entrenamiento.

Desarrollo

Gracias al aprendizaje profundo y otras áreas, es posible procesar grandes cantidades de datos como son texto, imágenes y videos. Algunos modelos son entrenados usando bases de datos de gran tamaño dando lugar a los conocidos LLM's.

Los modelos de inteligencia artificial generativa son capaces de generar texto, imagen y un gran variedad de contenido, permiten optimizar tiempo y creatividad gracias a dicha capacidad generativa.

Terminología básica

Se puede encontrar cierta dificultad para distinguir aprendizaje maquina e inteligencia artificial, de alguna forma se podrían confundir, a continuación se dan algunas definiciones que ayuden en la tarea de comprensión

- Inteligencia artificial, es un campo de las ciencias computacionales centrado en la creación de agentes inteligentes que razonen, aprendan y actúen de forma autónoma.
- Aprendizaje maquina, es un sub-campo de la IA que se enfoca en el desarrollo de algoritmos que permitan a la computadoras aprender de los datos.
- Aprendizaje profundo, son sistemas que mediante el uso de redes neuronales, compuestas de capas y modelos de aprendizaje maquina, aprenden de los datos de entrada.
- Modelos generativos, son un tipo de aprendizaje maquina que generan nuevo contenido basados en patrones aprendidos de los datos de entrada.
- Modelos de lenguaje, son modelos estadísticos usados para predecir palabras en una secuencia de lenguaje natural. Suelen usar aprendizaje profundo y ser entrenados en conjuntos de datos grandes, siendo denominados modelos largos de lenguaje o LLM's por sus siglas en ingles.

Los modelos generativos suelen ser muy buenos generando datos que imitan la información usada para su entrenamiento. Son capaces de desempeñarse para una gran variedad de tareas como texto a texto, texto a imagen, texto a audio, texto a video texto a voz, voz a texto, imagen a texto, imagen a imagen, texto a código, video a audio.

Resumen

Como se logra apreciar la inteligencia artificial es versátil, las definiciones que se dan sirven para ubicarse un poco en el mundo de la inteligencia artificial, para poder llegar a comprender de forma más clara las grandes aplicaciones que tiene y como se relaciona con otros conocimientos similares. Los datos parecen ser un tema poco relacionado, pero como ya se menciono son la fuente de inteligencia de los sistemas, pues son usados para entrenar a las inteligencias artificiales, la cantidad producida en la actividad humana diaria es enorme y deja rastro por internet.

Conclusiones

Cuando se introduce el tema de los datos en relación con IA, se pueden encontrar dificultades como el buen uso y tratamiento de los mismos, las regulaciones mundiales, o falta de regulaciones nacionales. Para fines prácticos es bueno recurrir a la ley nacional, o seguir algún acuerdo mundial.

Una definición común para dato personal es la que considera datos personales cualquier información que pueda llevar a identificar a una persona de forma directa o indirecta. A saber, mi dirección que sale en el id es una forma de ubicarme directamente. Por otro lado cuando

tomo una fotografía, el teléfono guarda la ubicación y fecha de la toma, incluso algunos guardan datos faciales, muchos datos de tal clase se les dice metadatos, o datos de los datos. En particular los metadatos permiten relacionar información de forma indirecta a una persona. Con eso en mente, es bueno comprender lo valioso que son los datos pero no olvidar lo delicado que es su tratamiento, permiten crear aplicaciones de un gran valor social pero a la vez podrían perjudicar a personas o infligir su privacidad.