

Checklist: Mantenimiento de las bases de datos para la IA

Bienvenido a la era de la inteligencia artificial (IA), donde la forma de hacer negocios depende de tecnologías intensivas en datos como machine learning y deep learning. Para aprovechar estas nuevas herramientas de IA, debe asegurarse de que las bases de datos de su organización estén en orden.

Revise el checklist para comenzar su camino correcto en las dos fases clave del mantenimiento de datos: entrenamiento e inferencia.

Siga estos pasos para ayudarlo a convertirse en un maestro de IA.

Para obtener más información, consulte este informe de IDC:

[*Acelere y operacionalice las implementaciones de IA utilizando infraestructura optimizada por IA.*](#)

Entrenamiento

En el aspecto de la preparación para la IA, desarrollará algoritmos para comprender un conjunto de datos. Su principal preocupación será recopilar datos existentes y utilizar la IA para aprender una nueva capacidad.

- Averigüe el problema comercial específico que querrá resolver utilizando IA (comience con proyectos más pequeños para ayudarlo a aprender)
- Localice los datos que pueden resolver ese problema de fuentes relevantes (lo más probable es que no todos estén ubicados en un solo lugar)
- Prepare sus datos con etiquetas de metadatos para reducir significativamente el tiempo requerido para encontrar datos pertinentes
- Asegúrese de que sus datos estén sincronizados y vinculados correctamente en todos los conjuntos de datos que utilizará (incluyendo la sincronización del tiempo)
- Marque todos los datos confidenciales y otros datos privados para asegurarse de mantenerlos absolutamente seguros y de cumplir con todas las normas y regulaciones apropiadas (el proceso de etiquetado de metadatos puede ayudar en este paso)
- Elija el entorno de desarrollo adecuado para el tipo de datos que está utilizando y la forma en que se formateará (es decir, imágenes, video, texto y audio, cada uno de ellos suele tener un tipo de entorno)
- Extraiga conjuntos de datos de su repositorio y llévelos a su entorno de desarrollo
- Divida sus datos en dos grupos para ayudarlo a mejorar el proceso de desarrollo de su modelo (mantenga un conjunto en una carpeta llamada "Capacitar" y otro conjunto en una carpeta llamada "Probar")
- Mantenga la trazabilidad de los datos haciendo un seguimiento de dónde / de qué fuente provienen sus datos (considere usar herramientas que puedan ayudar a automatizar el proceso)
- Realice tareas básicas de higiene de datos para preparar los datos para construir un modelo (por ejemplo, incluir el ingreso de datos faltantes y eliminar entradas nulas)
- Use una muestra de subconjunto de datos para los cuales ya conoce la respuesta a la actividad de predicción (esto se llama un "conjunto de entrenamiento") e identifique todos los pasos de preprocesamiento necesarios para preparar los datos para hacer una predicción
- Use su conocimiento de este conjunto de entrenamiento para calcular puntajes de precisión que pueden darle confianza para aplicar el mismo modelo a nuevos datos para los cuales el modelo nunca ha sido entrenado explícitamente

Inferencia

Una vez que haya desarrollado un modelo que funcione para resolver su problema comercial, pasará de la entrenamiento a la inferencia. En esta fase, está tomando ese modelo exitoso y aplicándolo a nuevos datos, lo que también requiere un mantenimiento de datos continuo.

- Ubique su modelo de IA cerca de sus datos para reducir la latencia, reducir los requisitos de ancho de banda y mejorar el rendimiento general del modelo
- Desarrolle un proceso eficiente de canalización de datos y aplique el etiquetado de metadatos a sus datos a medida que ingresen, de modo que se puedan recopilar nuevos datos y utilizarlos para mejorar el modelo en el futuro
- Etiquete los datos de una manera que esté vinculada y sincronizada (por ejemplo, si los datos están secuenciados en el tiempo, puede sincronizar los conjuntos de datos o vincularlos seleccionando un campo, como el nombre del cliente, en todos los datos que ingresan)
- Desarrolle un plan de almacenamiento de ciclo de vida de datos a largo plazo para saber cómo administrará el volumen y la velocidad de los datos a medida que ingresan y los archiva
- Considere contratar a un Chief Data Officer (CDO) para mantener la gestión de datos de su organización para futuros proyectos de inteligencia artificial, deep learning y otros proyectos basados en datos

