

Configuración de ambiente de desarrollo

Herramientas necesarias para el desafío

Nicolás Chirino



The better the question. The better the answer. The better the world works.



Shape the future
with confidence

1

- Ambiente de desarrollo

2

- API Open AI

3

- Disponibilidad de datasets

4

- Consejos y recomendaciones

Capítulo 1

Ambiente de Desarrollo

Google Colaboratory

A cada equipo se les asignará una cuenta de Google Colab con 200 unidades de compute disponible.

Horas de uso aproximadas con 200 unidades

Tipo de Hardware	Consumo (Unidades por Hora)	Horas Totales con 200 Unidades	Ideal para...
GPU T4	~ 2 unidades / hora	~ 100 horas	Tareas de entrenamiento estándar, inferencia y desarrollo general. Es la más eficiente en costo.
GPU L4	~ 4 unidades / hora	~ 50 horas	Un punto intermedio, más nueva y eficiente que la T4 para ciertas tareas.
GPU V100	~ 5 unidades / hora	~ 40 horas	Entrenamiento de modelos más complejos y pesados que requieren más memoria y velocidad que la T4.
GPU A100	~ 13 unidades / hora	~ 15 horas	El máximo rendimiento para entrenar modelos muy grandes (LLMs, modelos de difusión de alta resolución) en el menor tiempo posible.
CPU (Sin GPU)	~ 2 unidades / hora	~ 100 horas	Tareas que no requieren aceleración gráfica, como manipulación de datos con Pandas o Scikit-learn.

API Open AI

Modelos OpenAI

Modelos de Lenguaje y Visión (VLM)

Se les asignará una API KEY para acceder al endpoint del modelo GPT-4o mini. Esta clave les permitirá realizar consultas tanto de texto como de imágenes a través de la plataforma.

Modelos: gpt4-o-mini

Cantidad de tokens por equipo: ~100 MM.

Cantidad de imágenes a procesar: ~2000.

Cuidado con procesar videos.

- **Cantidad de videos de 320x240 de 9 minutos a procesar:** ~60

Con esto podrán realizar:

- Modelos de texto.
- Modelos de texto + imagen.
- Se les entregaran 7 USD antes de la hackatón para que puedan practicar.
- Se les entregaran 8 USD el día de la hackatón para realizar sus desarrollos.



Demo

Disponibilidad de dataset

Bases de datos

kaggle



- Pueden utilizar **cualquier base de datos que sea de dominio público**. Si desean emplear datos propios o de una fuente privada, deberán compartirlos y hacerlos accesibles para todos los demás participantes para garantizar una competencia justa.
- La Municipalidad de Las Condes compartió un set de datos públicos que podrían ser útiles, este se encuentra disponible en Google Drive y se los compartimos por correo.



Repositorio de recursos

Les dispondremos un repositorio donde deberán indicar los recursos a utilizar. Debe tener la siguiente estructura:

1. Integrantes del equipo
2. Mentor
3. Bases de datos a utilizar.
4. Modelos de código abierto a utilizar.
5. Códigos o recursos extra.
6. Principales librerías.

Esta información se debe subir como tipo markdown, haciendo un pull request en el repositorio **hasta el domingo 28 de septiembre**.

Integrantes del Equipo

- Alondra Araya
- Ignacio Barrera
- Martín Campos

Consultor Senior de EY

- Thomas Mayen

Bases de Datos de Dominio Público

A continuación se presentan algunos de los conjuntos de datos públicos que se utilizarán:

- [Roboflow Universe](#)
- [Railway Surveillance](#)
- [CCTV Curation Dataset](#)
- [Urban Night Back](#)
- [Datasets creados por nosotros](#)

Otros dataset de roboflow:

- <https://universe.roboflow.com/master-final-dataset-curation-v1a/cctv-curation-dataset-1-hhibk>
- <https://universe.roboflow.com/dataset-uutxr/cctv-naxyo>
- <https://universe.roboflow.com/binus-university-y2qnm/cctv-oofg7-fqsjq>

Demo

Consejos y recomendaciones

Consejos y recomendaciones

1. Ocupen sus cuentas personales de Google Colab en un comienzo.
2. Solo procesen videos cuando estén seguros de su desarrollo.
3. Adelanten la presentación del pitch lo más posible.
4. No se enfoquen en el “mejor modelo”, sino en cómo entregan valor.
5. Durante el pitch tendrán 3 minutos para presentar, piensen en cómo aprovecharlos!
6. Diviértanse.
7. Tomen agua.



Configuración de ambiente de desarrollo

Herramientas necesarias para el desafío

Nicolás Chirino



The better the question. The better the answer. The better the world works.



Shape the future
with confidence