# Todo el proceso (parámetros, tiempos, errores)

### 1. PARÁMETROS DEL ENTRENAMIENTO

* **Modelo base utilizado:**
  + Claude (vía API)
  + LLaMA
  + DeepSeek
* **Codificación del dataset final:**
  + **UTF-8**
* **Herramientas utilizadas:**
  + Google Colab (entorno principal)
  + AWS (entrenamiento avanzado)
  + Hugging Face Transformers
  + LangChain
  + Pinecone / ChromaDB (almacenamiento vectorial)
  + Sentence Transformers
  + NumPy, Scikit-learn, Seaborn, Plotly
* **Validación del modelo:**
  + Comparación antes y después del fine-tuning
  + Verificación semántica

**2. TIEMPOS POR FASE DEL PROYECTO**

|  |  |
| --- | --- |
| **Fase** | **Duración estimada** |
| Planeación y organización | \_\_1\_\_semana |
| Recolección y limpieza de datos | \_\_2\_\_días |
| Diseño del pipeline e interfaz | \_\_1\_\_días |
| Entrenamiento y fine-tuning | \_\_4\_\_días |
| Pruebas, validación y testing | \_\_2\_\_ días |
| Documentación final | \_\_5\_\_semanas |

**ERRORES DETECTADOS DURANTE EL PROYECTO**

* Respuestas demasiado generales sin contexto suficiente
* Cambios en respuestas comparativas
* Problemas al interpretar preguntas con errores ortográficos
* Sesgo hacia jugadores/equipos populares por datos desbalanceados
* Fechas mal formateadas