

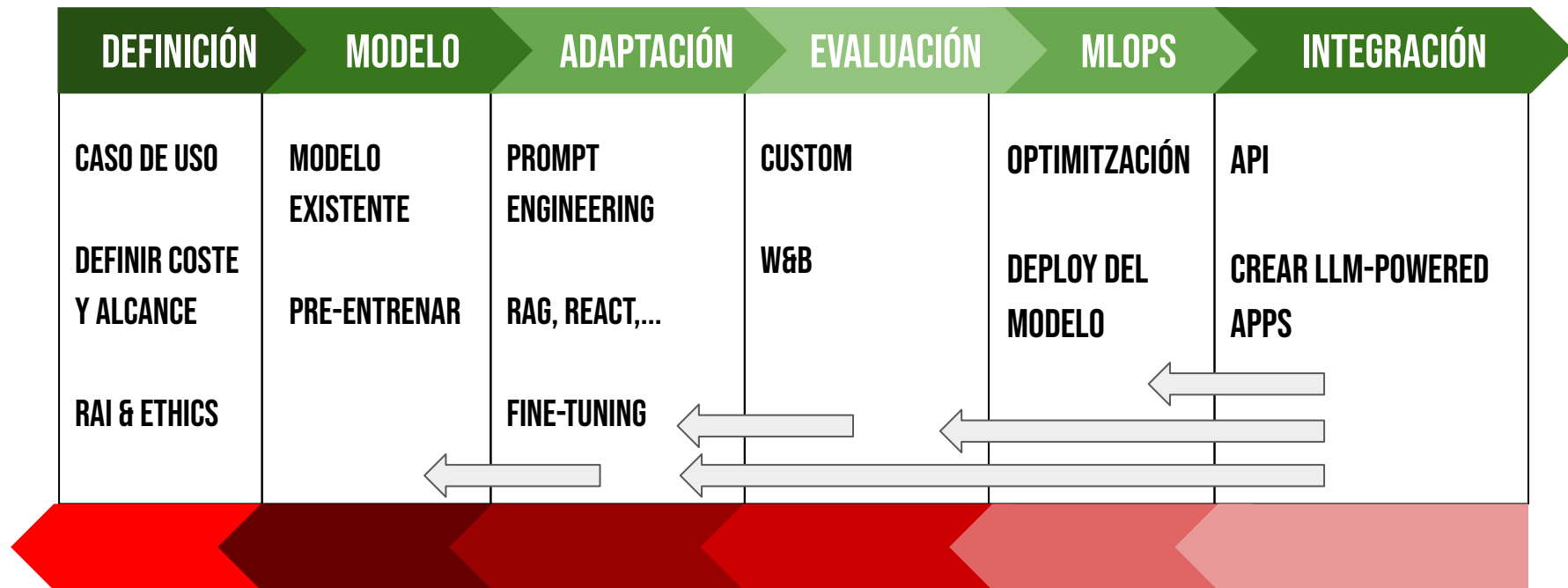
# LLM Engineering Práctica



**KEEPCODING**

Tech School

# GenAI Project Lifecycle





# ¿Que espero?

- Busquéis un problema a resolver.
- Buscar, crear, adaptar vuestros datos.
- Buscar y/o entrenar un modelo para resolver vuestro problema.
- Sed ingenieros.
- Crear una PoC o PoV de vuestro modelo.
- Podéis demostrar que funciona o que no.
- Compartid vuestros resultados con la comunidad (Hugging Face).
- Compartid vuestros resultados con LinkedIn! (Un post en LinkedIn).

# Ejemplos

- Necesito un modelo que sea capaz de contestar como si fuera un chatbot de asistencia al cliente.
- Quiero un modelo super experto en XXX
- Necesito conseguir que el modelo XXX sea lo menos tóxico posible
- Necesito encontrar cual es el mejor modelo en resolver XXX tarea
- Tengo una base de conocimiento enorme sobre XXX y necesito encontrar la mejor solución para vectorizarla
- Tengo que construir un agente y tengo que conseguir que no alucine ni se le pueda hacer inyecciones
- ...

# Mi ejemplo

- Quiero demostrar que el modelo Flor es un buen modelo para un Chatbot de asistencia al cliente
- Voy a partir del modelo fundacional y lo voy a finetunear para que responda mejor como un modelo de asistencia al cliente
- Necesito un dataset con Q&A en catalan
- Voy a finetunear una versión del modelo cuantificado Flor 1.3B
- Para demostrar que funciona lo voy a testear con un RAG sencillo.

# Mi ejemplo

- <https://colab.research.google.com/drive/1MkqKv7-OLGZHVOsypwZBL6k--aEtQbi?usp=sharing>
- <https://colab.research.google.com/drive/1DGtQVtw6tHqeDsMzbZ9sW2YUS0XoV3MM?usp=sharing>
- <https://colab.research.google.com/drive/1RdzS4piZBHlxHpsTsax82tfTj7K9lR5-?usp=sharing>
- <https://colab.research.google.com/drive/1jkRpz7VjdJsW4GVJ2DZu35lTQ0oPkro9?usp=sharing>



# Entregables

- Notebooks al Github.
- Una pequeña presentación explicando el problema, como buskais resolverlo y las conclusiones.
- Los modelos/datasets/spaces que subais a Hugging Face
- Una publicación a Linkedin con todo lo trabajado (opcional)
- Esto último de Linkedin!



# Bonus



- <https://docs.streamlit.io/develop/tutorials/llms/build-conversational-apps>





# KEEPCODING

## Tech School

Madrid | Barcelona | Bogotá

Eric Risco de la Torre

[erisco@icloud.com](mailto:erisco@icloud.com)

<https://www.linkedin.com/in/erisco-and/>