

HF Inference API + FastAPI + Streamlit — Chatbot

This is a minimal template for building a chatbot using Hugging Face's serverless Inference API, a FastAPI backend, and a Streamlit chat UI.

Prerequisites

- Python 3.10+
- A Hugging Face account and API token

Setup

See the per-folder READMEs in this document, or jump straight to **Run Locally** above.

Deploy

- Backend: Docker + any container host (Railway, Fly.io, Azure, etc.)
- Frontend: Streamlit Community Cloud (remember to set `BACKEND_URL`), or containerize alongside backend with CORS.