# Deployment Options for a Solo Developer

This brief compares three Python‑based frameworks—**Gradio**, **Streamlit** and **Chainlit**—for building chat‑style demos, and explores hosting options (Hugging Face Spaces, a small cloud VM and containerised self‑hosting). Each option is evaluated against decision criteria such as ease of development, authentication support, file‑size limits, cold‑start behaviour and cost. All factual statements link back to the cited sources.

## Framework comparison

| Framework | Strengths / Use cases | Authentication & sharing | Limitations | Key sources |
| --- | --- | --- | --- | --- |
| **Gradio 5.x/4.x** | Open‑source Python library that lets you build ML demos with just a few lines of code[[1]](https://pypi.org/project/gradio/#:~:text=Gradio%3A%20Build%20Machine%20Learning%20Web,Apps%20%E2%80%94%20in%20Python). The gr.ChatInterface class creates a fully‑functional chatbot UI from a single function[[2]](https://pypi.org/project/gradio/#:~:text=). Developers can run the app locally or in a notebook, and enable **hot‑reload** and **vibe mode** for rapid iteration[[3]](https://pypi.org/project/gradio/#:~:text=,in%20the%20%2041%20Hot). Built‑in sharing produces a public URL by passing share=True to .launch()[[4]](https://pypi.org/project/gradio/#:~:text=What%20good%20is%20a%20beautiful,the%20last%20line%20as%20follows). | No code is required beyond Python; sharing automatically spins up a public demo with a URL. Gradio apps can be hosted for free on **Hugging Face Spaces**, which is the most popular place to host Gradio applications[[5]](https://pypi.org/project/gradio/#:~:text=you%20build%20machine%20learning%20applications%2C,parts%20of%20the%20Gradio%20ecosystem). | For more customised UIs, developers must use the lower‑level gr.Blocks. ChatInterface is still evolving; recent changelog entries show numerous bug fixes around streaming and view API issues, but no long‑term stability guarantee[[6]](https://www.gradio.app/changelog). | Gradio documentation and PyPI page[[1]](https://pypi.org/project/gradio/#:~:text=Gradio%3A%20Build%20Machine%20Learning%20Web,Apps%20%E2%80%94%20in%20Python)[[2]](https://pypi.org/project/gradio/#:~:text=)[[5]](https://pypi.org/project/gradio/#:~:text=you%20build%20machine%20learning%20applications%2C,parts%20of%20the%20Gradio%20ecosystem). |
| **Streamlit** | Open‑source app framework for turning Python scripts into web apps in minutes. The official site emphasises that you can convert data scripts into shareable web apps using pure Python with **no front‑end experience required**[[7]](https://streamlit.io/#:~:text=A%20faster%20way%20to%20build,and%20share%20data%20apps). Streamlit Cloud enables public sharing for free and integrates with GitHub; the pro tier (Snowflake) supports unlimited private apps[[8]](https://streamlit.io/#:~:text=Or%20you%20can%20skip%20local,installation%20altogether). Building an app is as simple as writing a script and running streamlit hello[[9]](https://streamlit.io/#:~:text=Get%20started%20in%20under%20a,minute). | The Community Cloud automatically deploys from your GitHub repo. An app inherits the repo’s privacy: if the repo is private, the app is private; only one private app is allowed at a time[[10]](https://docs.streamlit.io/deploy/streamlit-community-cloud/share-your-app#:~:text=Your%20app%20is%20now%20live,from%20the%20app%20settings%20menu). Users can toggle between public/private from the app settings[[10]](https://docs.streamlit.io/deploy/streamlit-community-cloud/share-your-app#:~:text=Your%20app%20is%20now%20live,from%20the%20app%20settings%20menu). | Only supports Python; complex asynchronous tasks may need careful handling. Free tier requires public repositories; only one private app is allowed[[10]](https://docs.streamlit.io/deploy/streamlit-community-cloud/share-your-app#:~:text=Your%20app%20is%20now%20live,from%20the%20app%20settings%20menu). There may be cold‑start delays when apps sleep after inactivity [Unverified]. | Streamlit homepage and docs[[7]](https://streamlit.io/#:~:text=A%20faster%20way%20to%20build,and%20share%20data%20apps)[[8]](https://streamlit.io/#:~:text=Or%20you%20can%20skip%20local,installation%20altogether)[[10]](https://docs.streamlit.io/deploy/streamlit-community-cloud/share-your-app#:~:text=Your%20app%20is%20now%20live,from%20the%20app%20settings%20menu). |
| **Chainlit** | Open‑source Python package for building **production‑ready conversational AI**[[11]](https://docs.chainlit.io/get-started/overview#:~:text=Overview). Key features include the ability to **build fast** (few lines of code), integrate with corporate authentication, collect and monitor user data (data persistence), visualise multi‑step reasoning and support multi‑platform deployment[[11]](https://docs.chainlit.io/get-started/overview#:~:text=Overview). Chainlit apps can run as a web UI on port 8000, be embedded as a copilot in another site, integrated into custom React front‑ends or deployed to Teams, Slack or Discord[[12]](https://docs.chainlit.io/deploy/overview#:~:text=A%20Chainlit%20application%20can%20be,assistant%20logic%20once%2C%20use%20everywhere). | Provides built‑in authentication methods (password, header, OAuth) and supports hosting on various platforms. The deployment guide lists multiple self‑hosting options including AWS, Azure, Google Cloud Run, Replit, Fly.io and Hugging Face Spaces[[13]](https://docs.chainlit.io/deploy/overview#:~:text=After%20you%E2%80%99ve%20successfully%20set%20up,hosting%20your%20Chainlit%20app). Because Chainlit uses WebSockets, hosting services must support sticky sessions and CORS must be configured properly[[14]](https://docs.chainlit.io/deploy/overview#:~:text=By%20default%2C%20the%20Chainlit%20server,to%20your%20chainlit%20command). | Requires WebSocket‑friendly hosting. Developers need to configure --root-path when serving from a sub‑path and enable sticky sessions in load balancers[[14]](https://docs.chainlit.io/deploy/overview#:~:text=By%20default%2C%20the%20Chainlit%20server,to%20your%20chainlit%20command). No built‑in public URL generation; hosting must be arranged separately. | Chainlit docs[[11]](https://docs.chainlit.io/get-started/overview#:~:text=Overview)[[12]](https://docs.chainlit.io/deploy/overview#:~:text=A%20Chainlit%20application%20can%20be,assistant%20logic%20once%2C%20use%20everywhere)[[14]](https://docs.chainlit.io/deploy/overview#:~:text=By%20default%2C%20the%20Chainlit%20server,to%20your%20chainlit%20command). |

## Hosting options

| Hosting option | Pros / recommended scenarios | Limits / considerations | Cost (approx.) |
| --- | --- | --- | --- |
| **Hugging Face Spaces** | Provides a ready‑made platform for hosting Gradio, Streamlit or Docker apps[[15]](https://huggingface.co/spaces/launch#:~:text=Build%20it%20your%20way). Spaces offer git‑based version control, secrets management and the ability to upgrade hardware (Zero GPU)[[16]](https://huggingface.co/spaces/launch#:~:text=Is%20there%20a%20limit%20of,many%20Spaces%20I%20can%20have). Each Space runs in an environment with up to **16 GB RAM and 8 CPU cores**[[17]](https://huggingface.co/spaces/launch#:~:text=Each%20environment%20is%20currently%20limited,RAM%20and%208%20CPU%20cores). Developers can create unlimited Spaces and manage secrets without hard‑coding tokens[[18]](https://huggingface.co/spaces/launch#:~:text=Is%20there%20a%20limit%20of,many%20Spaces%20I%20can%20have). | Free tier uses a basic CPU; heavy models require upgrading to GPU. Spaces must remain active—long periods of inactivity may lead to cold‑starts. Storage is non‑persistent by default, so large files must be stored elsewhere [Unverified]. | Free for CPU instances; GPU upgrade costs start at \$0.60–\$1.00 per hour (approx.) [Unverified]. |
| **Small VM (e.g., DigitalOcean Droplet)** | Gives full control over OS, dependencies and privacy. A **Basic Droplet** with 1 vCPU and 1 GB RAM costs about **\$6 per month** and includes 25 GB SSD and 1 TB transfer[[19]](https://onedollarvps.com/pricing/digitalocean-pricing.html#:~:text=Droplets%20); a 0.5 GB RAM plan is \$4 per month[[20]](https://onedollarvps.com/pricing/digitalocean-pricing.html#:~:text=Plan%20vCPU%20RAM%20SSD%20Transfer,2gb%201%202GB%2050GB%202TB%2412). Suitable for self‑hosting a Python server or containerised app. | Requires DevOps skills: user must provision the VM, configure firewall/SSL, deploy via Docker or systemd, and monitor usage. Cold‑start times depend on VM boot and app start; uptime is constant (no sleeping). | \$4–\$12 per month for small instances[[21]](https://onedollarvps.com/pricing/digitalocean-pricing.html#:~:text=Plan%20vCPU%20RAM%20SSD%20Transfer,16gb%208%2016GB%20320GB%206TB%2496). Additional cost for domain, backups and bandwidth. |
| **Container (self‑hosted via Docker/Kubernetes)** | Enables local or on‑premises deployment. A Chainlit or Gradio app can run as a Docker container and be deployed to platforms like AWS ECS, Google Cloud Run or Kubernetes clusters. Chainlit docs note that hosting services must support WebSockets and may require sticky sessions or root‑path configuration[[14]](https://docs.chainlit.io/deploy/overview#:~:text=By%20default%2C%20the%20Chainlit%20server,to%20your%20chainlit%20command). | Requires container registry and orchestrator; complexity scales with traffic. Not ideal for beginners; heavy for simple demos. Cost depends on underlying compute (e.g., cloud VM or Kubernetes cluster). | Highly variable; running on a personal server is free besides electricity; managed services bill per CPU/GB (e.g., Google Cloud Run charges for CPU/memory by the second) [Unverified]. |

## Decision criteria and recommendations

1. **Ease of development:** For a solo developer seeking a quick demo, **Gradio** offers the simplest path. Its ChatInterface creates a chat UI with minimal code and sharing a demo is as easy as passing share=True[[4]](https://pypi.org/project/gradio/#:~:text=What%20good%20is%20a%20beautiful,the%20last%20line%20as%20follows). Streamlit also allows rapid scripting of data‑centric apps[[7]](https://streamlit.io/#:~:text=A%20faster%20way%20to%20build,and%20share%20data%20apps), but lacks a built‑in chatbot abstraction. Chainlit is best when you need advanced conversational features (multi‑step reasoning, user session storage)[[11]](https://docs.chainlit.io/get-started/overview#:~:text=Overview).
2. **Authentication and privacy:** If you need private access control without external infrastructure, Chainlit provides authentication integrations (password, header, OAuth)[[11]](https://docs.chainlit.io/get-started/overview#:~:text=Overview). Streamlit Community Cloud inherits GitHub permissions and allows only one private app on the free tier[[10]](https://docs.streamlit.io/deploy/streamlit-community-cloud/share-your-app#:~:text=Your%20app%20is%20now%20live,from%20the%20app%20settings%20menu). Gradio’s free sharing is public; to protect private documents you should self‑host or use a VM.
3. **Hosting convenience:** For hassle‑free deployment and minimal cost, **Hugging Face Spaces** is ideal. It supports Gradio, Streamlit and Docker, offers secrets management, and lets you upgrade to GPU when needed[[15]](https://huggingface.co/spaces/launch#:~:text=Build%20it%20your%20way)[[16]](https://huggingface.co/spaces/launch#:~:text=Is%20there%20a%20limit%20of,many%20Spaces%20I%20can%20have). However, memory and CPU are capped at 16 GB and 8 cores[[17]](https://huggingface.co/spaces/launch#:~:text=Each%20environment%20is%20currently%20limited,RAM%20and%208%20CPU%20cores). If you need persistent storage or complete control, rent a small VM (e.g., \$6/month for 1 GB RAM[[22]](https://onedollarvps.com/pricing/digitalocean-pricing.html#:~:text=Plan%20vCPU%20RAM%20SSD%20Transfer,4gb%202%204GB%2080GB%204TB%2424)). Chainlit’s WebSocket requirements mean that some managed platforms may require sticky sessions and CORS configuration[[14]](https://docs.chainlit.io/deploy/overview#:~:text=By%20default%2C%20the%20Chainlit%20server,to%20your%20chainlit%20command).
4. **Cold start behaviour:** On free hosting platforms (Spaces, Streamlit Cloud) your app may sleep after inactivity, leading to 30‑60 second cold‑start delays [Unverified]. Running on a dedicated VM or container avoids sleeping and yields consistent latency, at the cost of continuous billing.
5. **Cost:** Free tiers (Spaces, Streamlit Community Cloud) are well‑suited to prototypes. Upgrading to GPU or private instances introduces monthly costs. Self‑hosting on a small VM starts at \$4–\$6 per month[[21]](https://onedollarvps.com/pricing/digitalocean-pricing.html#:~:text=Plan%20vCPU%20RAM%20SSD%20Transfer,16gb%208%2016GB%20320GB%206TB%2496) and scales linearly with resources. Containerised deployments on cloud platforms have variable cost depending on CPU and memory usage [Unverified].

## Recommendation for a solo developer

Begin with **Gradio ChatInterface** to build a prototype quickly. Use **Hugging Face Spaces** for free hosting; you will get a shareable URL and secrets management[[5]](https://pypi.org/project/gradio/#:~:text=you%20build%20machine%20learning%20applications%2C,parts%20of%20the%20Gradio%20ecosystem)[[16]](https://huggingface.co/spaces/launch#:~:text=Is%20there%20a%20limit%20of,many%20Spaces%20I%20can%20have). As your app matures and requires custom authentication or more control, consider migrating to **Chainlit** and deploying on a **small VM** or container orchestration. Streamlit is an excellent alternative if your app is primarily data‑visualisation rather than conversational.

[[1]](https://pypi.org/project/gradio/#:~:text=Gradio%3A%20Build%20Machine%20Learning%20Web,Apps%20%E2%80%94%20in%20Python) [[2]](https://pypi.org/project/gradio/#:~:text=) [[3]](https://pypi.org/project/gradio/#:~:text=,in%20the%20%2041%20Hot) [[4]](https://pypi.org/project/gradio/#:~:text=What%20good%20is%20a%20beautiful,the%20last%20line%20as%20follows) [[5]](https://pypi.org/project/gradio/#:~:text=you%20build%20machine%20learning%20applications%2C,parts%20of%20the%20Gradio%20ecosystem) gradio · PyPI

<https://pypi.org/project/gradio/>

[[6]](https://www.gradio.app/changelog) Gradio Changelog

<https://www.gradio.app/changelog>

[[7]](https://streamlit.io/#:~:text=A%20faster%20way%20to%20build,and%20share%20data%20apps) [[8]](https://streamlit.io/#:~:text=Or%20you%20can%20skip%20local,installation%20altogether) [[9]](https://streamlit.io/#:~:text=Get%20started%20in%20under%20a,minute) Streamlit • A faster way to build and share data apps

<https://streamlit.io/>

[[10]](https://docs.streamlit.io/deploy/streamlit-community-cloud/share-your-app#:~:text=Your%20app%20is%20now%20live,from%20the%20app%20settings%20menu) Share your app - Streamlit Docs

<https://docs.streamlit.io/deploy/streamlit-community-cloud/share-your-app>

[[11]](https://docs.chainlit.io/get-started/overview#:~:text=Overview) Overview - Chainlit

<https://docs.chainlit.io/get-started/overview>

[[12]](https://docs.chainlit.io/deploy/overview#:~:text=A%20Chainlit%20application%20can%20be,assistant%20logic%20once%2C%20use%20everywhere) [[13]](https://docs.chainlit.io/deploy/overview#:~:text=After%20you%E2%80%99ve%20successfully%20set%20up,hosting%20your%20Chainlit%20app) [[14]](https://docs.chainlit.io/deploy/overview#:~:text=By%20default%2C%20the%20Chainlit%20server,to%20your%20chainlit%20command) Overview - Chainlit

<https://docs.chainlit.io/deploy/overview>

[[15]](https://huggingface.co/spaces/launch#:~:text=Build%20it%20your%20way) [[16]](https://huggingface.co/spaces/launch#:~:text=Is%20there%20a%20limit%20of,many%20Spaces%20I%20can%20have) [[17]](https://huggingface.co/spaces/launch#:~:text=Each%20environment%20is%20currently%20limited,RAM%20and%208%20CPU%20cores) [[18]](https://huggingface.co/spaces/launch#:~:text=Is%20there%20a%20limit%20of,many%20Spaces%20I%20can%20have) Spaces Launch – Hugging Face

<https://huggingface.co/spaces/launch>

[[19]](https://onedollarvps.com/pricing/digitalocean-pricing.html#:~:text=Droplets%20) [[20]](https://onedollarvps.com/pricing/digitalocean-pricing.html#:~:text=Plan%20vCPU%20RAM%20SSD%20Transfer,2gb%201%202GB%2050GB%202TB%2412) [[21]](https://onedollarvps.com/pricing/digitalocean-pricing.html#:~:text=Plan%20vCPU%20RAM%20SSD%20Transfer,16gb%208%2016GB%20320GB%206TB%2496) [[22]](https://onedollarvps.com/pricing/digitalocean-pricing.html#:~:text=Plan%20vCPU%20RAM%20SSD%20Transfer,4gb%202%204GB%2080GB%204TB%2424) DigitalOcean Pricing: VPS, GPU, Storage and More (2025)

<https://onedollarvps.com/pricing/digitalocean-pricing.html>