

# 1. Best Self-Hostable Tools for Character Consistency

**Stable Diffusion (SD)** – **Stability AI's Stable Diffusion** is the leading open-source image generation model for self-hosting. Unlike closed services (MidJourney, DALL-E) which lack fine-tuning, Stable Diffusion's open models can be run on your own GPU server and **customized to learn a specific character** <sup>1</sup>. The community around SD has created many free tools and methods for character control and consistency <sup>1</sup>. You can choose from SD model versions (v1.5, v2.1 or the newer SDXL for higher quality) – all are self-hostable.

**Local Deployment and UIs:** To run SD on your server, popular interfaces include **AUTOMATIC1111's Stable Diffusion Web UI** and **ComfyUI**. AUTOMATIC1111's web UI (available on GitHub) is user-friendly and supports extensions for training and ControlNet. It allows you to load models, generate images, and train embeddings/LoRAs through a browser. ComfyUI is a node-based interface for advanced workflows. Both are open-source and can be installed on a server with an NVIDIA GPU. Basic requirements are a GPU with at least ~6–8 GB VRAM for standard SD (more for training large models or high-res images). *(For example, SDXL models demand ~12 GB VRAM to train or use.)* Ensure you have the SD model weights (download from HuggingFace or official sources) and that your environment has PyTorch installed.

## DreamBooth Fine-Tuning (Full Model Fine-tune)

DreamBooth is a method to **fine-tune the entire model** to learn your character (or any subject). Originally by Google, it showed that with as few as 3–5 images of a subject, one can fine-tune a diffusion model to insert that subject into new scenes. DreamBooth incorporates the concept by training with a **unique identifier token + a class noun** (e.g. “a photo of *zyxOwl* owl”) so that the model learns your character “*zyxOwl*” as a specific instance of the class “owl”. After training, you use that token in prompts to get the character.

- *Training:* DreamBooth is the most **resource-intensive** method. You typically need a higher-end GPU (generally  $\geq 12$  GB VRAM recommended) to fine-tune the model weights. The process can be done with libraries like HuggingFace Diffusers or via extensions (there are DreamBooth extensions for AUTOMATIC1111, and also Google Colab notebooks available). You provide a set of images of your character (5–10 images from different angles/poses is great) and choose a **unique token** (a made-up word unlikely to appear in normal prompts) and an appropriate class word. The training will gradually adjust the model’s weights to reproduce your character in those images while trying to keep the model’s general ability (often a regularization technique is used – e.g., generating some “class images” of a generic owl to not forget how to draw owls in general). The output is a **new model checkpoint** (or diff file) typically several GB in size.