SYSTEM ARCHITECTURE & HARDWARE PREREQUISITES

Project: Local Generative AI Studio Deployment

Version: 2.0 (2026 Stack)

Technician Note: NVIDIA GPU Architecture is Mandatory.

1. EXECUTIVE SUMMARY

This installation transforms the client's workstation into a high-performance, offline inference engine. Unlike cloud software, performance is strictly tied to local hardware capabilities. To ensure the functionality of the **2026 Tech Stack** (Flux.2, Wan 2.1, LTX-2), the following specifications must be met.

2. HARDWARE COMPATIBILITY MATRIX

TIER 1: MINIMUM VIABLE (Drafting Station)

Capable of running Flux Schnell (Images) and LTX-2 (720p Video). Slower render times.

- **GPU:** NVIDIA RTX 3060 (12GB VRAM) or RTX 4060 Ti (16GB VRAM).
- **System RAM:** 32 GB DDR4.
- **CPU:** Intel Core i5 (12th Gen) or AMD Ryzen 5 (5000 Series).
- **Storage:** 500 GB free on NVMe SSD.

TIER 2: RECOMMENDED (Pro Creator)

Capable of running Flux.2 Pro, Wan 2.1 Video (1080p), and real-time Z-Image Turbo. Smooth workflow.

- **GPU:** NVIDIA RTX 4070 Ti Super (16GB) or RTX 3090 (24GB).
- **System RAM:** 64 GB DDR5 (Critical for Video Model loading).
- **CPU:** Intel Core i7 (13th/14th Gen) or AMD Ryzen 9 (7900X).
- **Storage:** 2 TB Gen4 NVMe SSD.

TIER 3: PRODUCTION STUDIO (Enterprise)

Capable of heavy Batch Processing, Training (LoRA), and 4K Video Generation.

- **GPU:** NVIDIA RTX 4090 (24GB) or RTX 6000 Ada.
- **System RAM:** 128 GB DDR5.
- **CPU:** Intel Core i9 (14900K) or AMD Ryzen 9 (7950X).
- **Storage:** 4 TB Gen4 NVMe + Secondary SSD for Dataset storage.

3. CRITICAL DEPENDENCIES

A. GPU Architecture (CUDA)

- **Supported:** NVIDIA RTX 3000 series, 4000 series, 5000 series (desktop).
- **Not Supported:** AMD Radeon, Intel Arc, Apple Silicon (M1/M2/M3), Integrated Graphics.
- **Driver:** Must be updated to Game Ready Driver 560.xx or higher.

B. Operating Environment

- **OS:** Windows 10 or Windows 11 (64-bit Pro recommended).
- **Permissions:** Local Administrator access required for installation of Python dependencies, Triton kernels, and C++ build tools.

C. Network (Installation Only)

- While the final system is 100% offline, a stable internet connection (min 50 Mbps) is required during the 3-5 hour installation window to download approximately **150GB - 200GB** of model weights and dependencies.

4. STORAGE ALLOCATION BREAKDOWN

The AI Stack requires significant space. Please ensure the C: or D: drive has contiguous free space.

- **Base Dependencies (Python/CUDA):** ~15 GB
- **Image Models (Flux.2/SDXL/Chroma):** ~60 GB
- **Video Models (Wan 2.1/LTX-2):** ~45 GB
- **Upscalers & ControlNets:** ~20 GB
- **Swap Space/Cache:** ~20 GB
- **TOTAL RECOMMENDED FREE SPACE: 200 GB**

Client Acknowledgement:

I confirm my hardware meets the criteria above. I understand that hardware below the "Minimum" spec will result in software instability.

Date: _____