

Aggressive Kühe auf der ArchivistaBox

Egg, 15. April: Ab sofort ist die Integration von ComfyUI für die ArchivistaBox AGPLv3 erhältlich. ComfyUI, noch nie davon gehört? Kein Problem, in diesem Blog wird aufgezeigt, was ComfyUI ist. Weiter wird anhand eines Beispiels mit «aggressiven Kühen» gezeigt, welche Resultate aktuell mit Open Source möglich sind und drittens wird am Ende die Installation auf der ArchivistaBox so erklärt, dass sie mühelos gelingt.



Was ist ComfyUI?

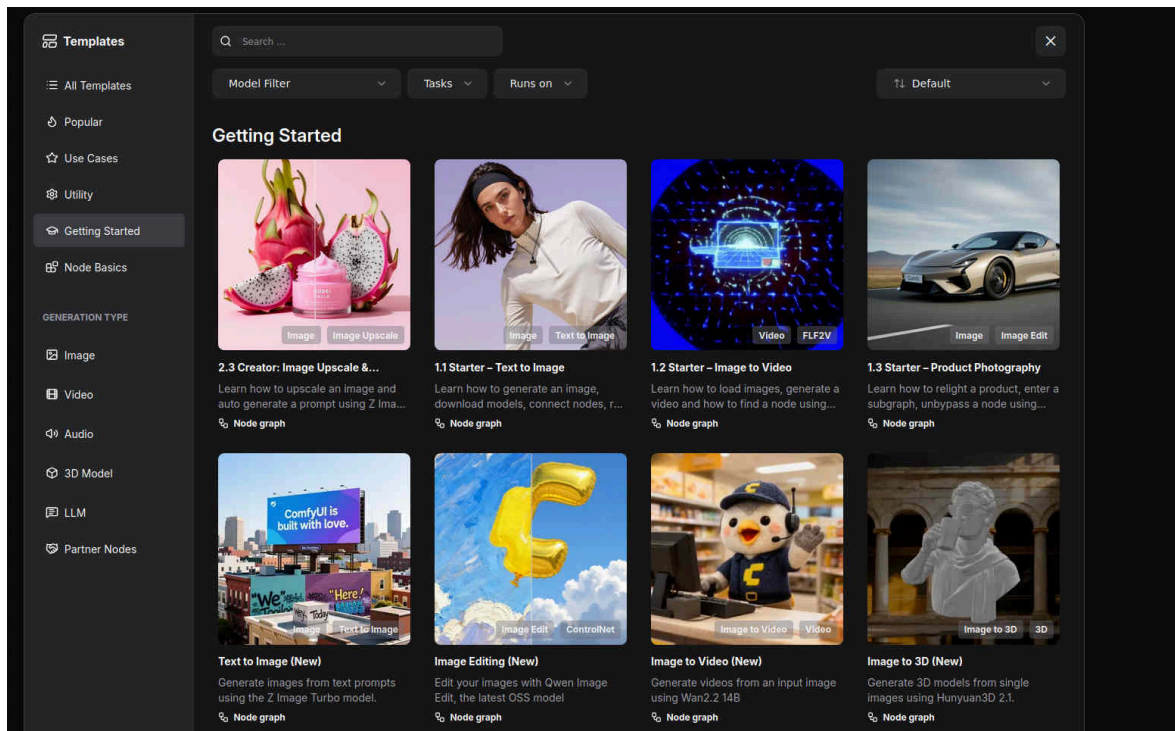
Wer mit KI-Modellen arbeiten möchte, benötigt dazu ein Programm, mit dem den Modellen gesagt wird, was zu tun ist. Ich möchte dies an einem Beispiel erläutern. für das Projekt <https://ch1291.ch> sollte ein kurzer Song bzw. Clip entstehen, um das Vrohaben vorzustellen. Zunächst hier der Link, um den Clip anzusehen:

Start und Songs für das Projekt

Dabei wurde ein Song über 90 Sekunden in den Sprachen Mundar, Hochdeutsch, Französisch, Italienisch und Englisch erstellt. Ebenso wurden für den Clip 21 Bilder generiert, um den Song bildlich im wahrsten Sinne des Wortes zu untermalen. In einer Sequenz geht es um das Konfliktpotential zwischen Mensch und Kühen beim Wandern. Dies wäre in der Realität etwas schwierig zu realisieren. Wer will schon Kühe provozieren, um eine gute Aufnahme zu erhalten? Ebenso wird wohl niemand die Kamera zücken, wenn die Kühe im Anmarsch sind.

Im Prinzip könnte über ein Bildbearbeitungsprogramm versucht werden, Kühe und ein Mensch in Szene zu setzen. Ist aber nicht ganz einfach und auch Profis dürften hier gefordert sein. Ein klassischer Fall für die KI.

Damit dies gelingt, braucht es a) eine gute Grafikkarte, b) ein Modell, das aus Texten Bilder erzeugt und c) ein Programm, das dem Modell sagt, was zu tun ist. Idealerweise arbeitet dieses Programm als Web-Applikation. Klar ist zudem, bei einem Modell, das lokal läuft, soll auch das Programm, um diesen Job zu erledigen, lokal laufen. Genau dies bietet ComfyUI. Einmal installiert, kann aus einer Vielzahl von Vorlagen der Job sehr einfach auf dem lokalen Rechner erledigt werden.



Vier Modelle versuchen ihr Glück

Die Anzahl der Modelle mag am Anfang bei ComfyUI etwas verwirrend sein. Letztlich ist es aber einfach so, dass es mittlerweile eine beachtliche Anzahl von freien KI-Modellen gibt, sozusagen ganz nach dem Motto, wer die Wahl hat, hat die Qual.

Vorweg sei gesagt, einfache Aufgaben gelingen mit sämtlichen Modellen.

Beispiel: «Erstelle ein Bild, das eine Blumenwiese zeigt». Schwieriger wird der Fall, wenn die Szene komplexer ist. In unserem Beispiel soll eine Szene mit vier aggressiven Kühen entstehen, bei welcher der Wanderer flüchtet:

Vier aggressive Kühe in den Alpen in der Nähe vom Mont-Blanc greifen älteren männlichen Wanderer mit Rucksack von hinten an, Wanderer und Kühe rennen in gleiche Richtung, es regnet in Strömen

Warum ist dies nicht so einfach für die KI? Erstens sollen es ja nicht einfach vier Kühe sein, sondern sie sollen aggressiv dreinschauen. Zweitens gesellt sich der Wanderer dazu, der Angst vor den Kühen hat und flüchtet. Und drittens bringt es ja nichts, wenn der Wanderer angstvoll zu den Kühen hinrennt, denn dies wäre ja eine Art Kamikaze-Aktion. Also, der Wanderer muss in die gleiche Richtung wie die Kühe rennen, und überdies vor und nicht hinter den Kühen. Die Kulisse, dass es in der Nähe des Mont-Blanc spielen soll, dient dazu da, herauszufinden, ob das Modell mit geographischen Begebenheiten klarkommt. Überdies wurde noch der Regen eingebaut, dann macht eine Flucht bekanntlich am meisten Spass ;-)

ZImage-Turbo: Der strauchelnde Wanderer



Es läuft da einiges schief. Die drei Kühe (es sollten an sich vier sein) erscheinen quasi als Klone, der Wanderer rennt in die falsche Richtung und die Beine sind schlicht falsch.

Angefügt sei hier, dass das Modell Z-Image-Turbo mit vergleichsweise kleinen Ressourcen klarkommt (8 bzw. 16 GB VRAM). Aber, das Resultat ist faktisch unbrauchbar. Wie bereits erwähnt, einfachere Dinge gelingen schon, einfach in dieser etwas komplexeren Szene ist Z-Image-Turbo so ziemlich überfordert.

Z-Image: Klar sind die Kühe aggressiv, der Wanderer provoziert



Das «normale» ZImage-Modell erfordert 32 GByte bei der Grafikkarte. Die Kühe sind nun perfekt aggressiv, nur rennt der Wanderer in die falsche Richtung, nämlich auf die Kühe zu. War nicht die Aufgabe, aber es lässt sich doch erahnen, dass es zu einem Kampf kommen könnte.

Qwen-Image-2512: Kühe und Wanderer in Einklang



Qwen-Image-2512 bedeutet zunächst einmal, dass das Modell vom Dezember 2025 datiert. Alle vier hier getesteten Modelle sind nicht älter als einige wenige Monate. Fakt ist, die Entwicklung bei den lokalen KI-Modellen ist längst nicht abgeschlossen. Offen gestanden waren Open Source Modelle vor ein zwei Jahren noch deutlich weniger an einem Punkt, dass sie lokal eingesetzt werden konnten.

Angefügt sei hier noch, dass Bild umfasst nur drei Kühe. Es ist davon auszugehen, dass das KI-Modell den Wanderer zu den Kühen «zählte». Aber immerhin, in die gleiche Richtung schauen sie schon mal. Weiter ist der Wanderer sehr dynamisch unterwegs. Als würde er mitten durch die «Herde» rennen...

Flux-Dev: Der «Testsieger» lässt nichts anbrennen



Die Kühe sind artig aggressiv, die Zahl stimmt, der Wanderer davor, und er rennt von den Kühen davon. Kurz und gut, für den ersten Versuch nicht schlecht. Es könnte «kritisiert» werden, dass die Kühe links und der Wanderer

recht im Bild sein müssten (oder umgekehrt). Aber, dass wir ja nicht die Aufgabenstellung.

In diesem Sinne ist Flux-Dev der Testsieger. Der Preis dafür ist einigermaßen hoch. Unter 32 GByte VRAM bei der Grafikkarte keine Chance. Und mit 32 GB VRAM auch nur lauffähig, wenn gewisse Teile bei der Verarbeitung im Hauptspeicher liegen. Allerdings führt dies nicht zu einer wesentlich verlängerten Rechenzeit. Insgesamt benötigt Flux-Dev bei 1920×1088 Pixeln (FullHD) in etwa zwei bis drei Minuten Rechenzeit, dies gemessen mit der AMD-Grafikkarte R9700.

Darum ist die AMD R9700 aktuell die perfekte Wahl

Wie oben aufgeführt, ist anspruchsvollere Bild-zu-Text-KI mit 16 GB Speicher bei der Grafikkarte aktuell kaum machbar. Bei 32 GByte sieht es so aus, dass die Preise bei NVidia recht hoch sind. 3000 Franken oder mehr alleine für die Grafikkarte, das ist nicht nichts. Die R9700 von AMD kostet aktuell 1200 Franken (beide Preise inkl. MwSt).

Dieser Preisunterschied ist dem geschuldet, dass NVidia der Platzhirsch darstellt. Wie einfach es ist, die Treiber bei AMD unter Windows zu aktivieren, dies wurde nicht getestet. Unter Linux kann ausgeführt werden, dass die Treiber-Installation im Prinzip nicht mehr allzu schwierig ist. Etwas schwieriger wird die Sache dann, wenn es darum geht ComfyUI zu installieren. Und noch anspruchsvoller wird es, wenn Flash Attention 2 aktiviert werden soll. Dieses Modul ist zwar «optional», bietet aber schnellere Abarbeitung an.

Einmal korrekt eingerichtet, bietet die Kombination zwischen RX 9070 XT bzw. R9700 sehr gute Ergebnisse. Modelle über 32 GByte können entweder durch Off-Loading über die CPU (funktioniert bei Flux-Dev sehr gut) oder aber über eine zweite Grafikkarte via PCI5-x16 realisiert werden.

Mit der R9700 wurden auch Modelle zur Generierung von Videos getestet. Besonders das aktuelle (März 2026) LTX2.3 bietet eine erstaunliche Qualität in FullHD (Upscaling auf 4K möglich) und Synchronisierung der Audio- wie Video-Spur. Will heißen, es können Dialoge in die Szenen eingebaut werden, dies mit recht perfektem Deutsch. Mit etwas Übung konnten auch Dialekte realisiert werden.

Angefügt sei hier, dass für gute Resultate oft einige Durchläufe notwendig sind. Im Unterschied zu Text-zu-Bild ist es bei Text-zu-Video ja so, dass die textliche Beschreibung quasi in ein Skript übersetzt werden muss. Ebenso sind pro Sekunde minimal 25 Bilder/Sekunde notwendig, bei LTX2.3 können aktuell bis 50 Bilder/Sekunde erstellt werden.

Cloud-Angebote nicht unbedingt preiswert

Wer mit geschlossenen Cloud-Angeboten arbeitet, wird schnell feststellen, dass Preis-Modelle, die bei 100 Franken pro Monat liegen (bzw. die entsprechenden Credits), nicht allzu weit reichen (pro Minute FullHD-Video sind es gerne um die 40 USD). Geschätzt wurden bislang auf der ArchivistaBox KI mit ComfyUI zwei bis drei Stunden Clips realisiert.

Dies hätte beim Platzhirsch die «Kleinigkeit» von ca. 7000 USD gekostet. Ohne eigene lokale KI wäre dies ein sehr teures Unterfangen. Und es mag sein, dass ohne eigene KI die Ritter nicht in derart vielen Facetten durch die Gegend marschiert wären. Der kurze Clip (mehr gibt die Bandbreite des aktuellen Providers nicht her) soll illustrieren, wie ein Resultat aussehen kann.

<https://archivista.ch/cms/wp-content/uploads/burg1.mp4>

Mit LTX2.3 sind Szenen mit 20 bis 30 Sekunden in FullHD bzw. 4K (Upscaling) möglich.

Die Audio-Synchronisation von Version 2.0 war noch nicht wirklich brauchbar, jene der neuen Version schon. Bei einem 5-Sekunden-Clip gibt Sprache wenig Sinn, zu gegebener Zeit wird es aber gerne Beispiele geben, welche gesprochene Szenen enthalten. Ferner finden sich im Netz zahlreiche Demos. Die dort gezeigte Qualität kann bestätigt werden.

Installation von ComfyUI auf der ArchivistaBox

Zunächst einmal ist die aktuelle ArchivistaBox 2026/III sowie das Modul ai.os notwendig. Dieser **Vorgang ist im letzten Blog bereits erklärt**. Weiter wird neu die Datei comfyui.os benötigt. Diese kann hier bezogen werden:

<https://archivista.ch/cms/comfyui>

Die Datei ,comfyui.os beziehen und unter **/home/archivista/data** entpacken:

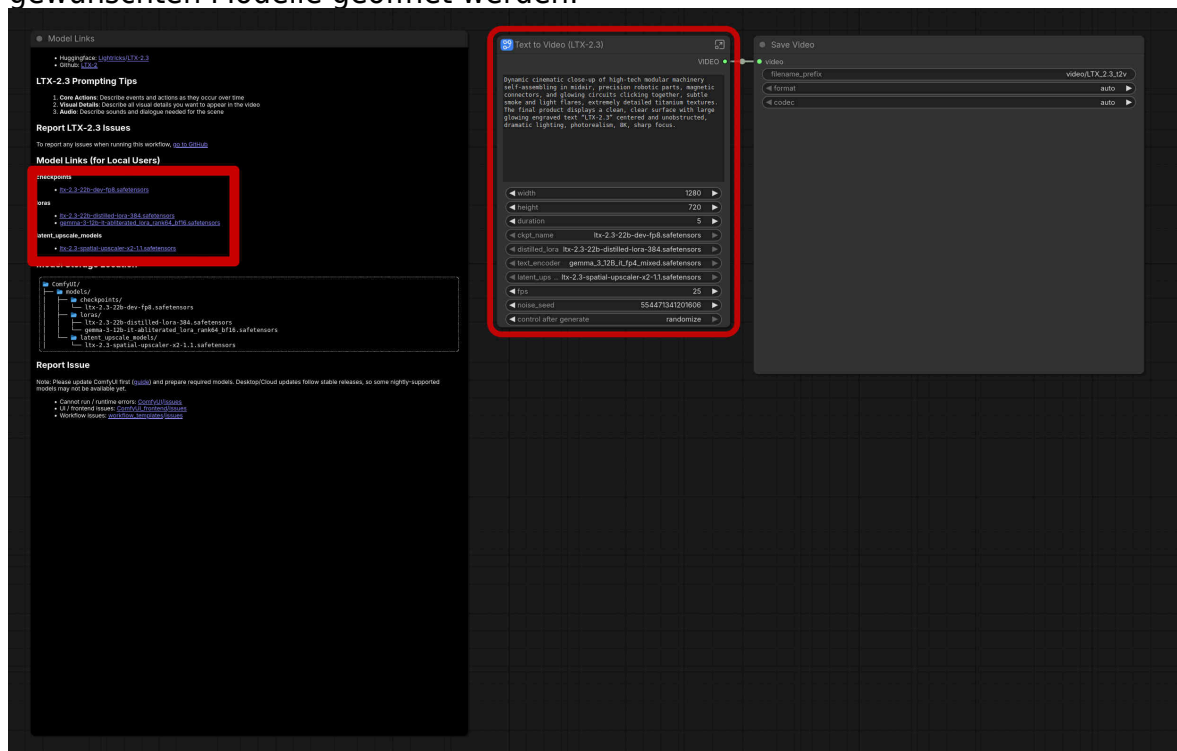
unsquashfs -d comfyui comfyui.os

Danach mit ,cd comfyui' in den Ordner wechseln und die virtuelle Umgebung öffnen:

source comfyui/bin/activate

Es erscheint danach der Prompt ,comfyui' in der Konsole. Jetzt mit cd ComfyUI in den entsprechenden Ordner wechseln. ComfyUI kann nun über ./comfyui-amd.sh gestartet werden.

ComfyUI wird ist danach über ,http://localhost:8189' bzw. die Adresse des Rechners unter dem Port 8189 erreicht. Das aktuelle Paket umfasst ca. 3 GByte. Ausgepackt sind es ca. 7 GByte. In ComfyUI können nun über Templates die gewünschten Modelle geöffnet werden:



Der rote Rahmen in der Mitte sagt, es fehlen Modelle. Links findet sich (hier mit dickem roten Rand umrahmt) die entsprechenden Modelle, die benötigt werden bzw. an welchen Ort sie abgelegt werden müssen. Am einfachsten ist die Installation, indem Ctrl+Enter gedrückt wird. Dabei wird immer der Job gestartet. ComfyUI meldet dann von alleine, welche Dateien fehlen. Ebenso kann dort dann direkt der Download gestartet werden.

Wichtig zu wissen ist, die Modelle werden nach /home/archivista/data heruntergeladen. Von dort müssen sie gemäss Hilfestellung im Template in den jeweiligen richtigen Ordner von ComfyUI verschoben werden. Danach erneut Ctrl+Enter drücken, einfach solange bis der Job läuft (grüner anstelle von rotem Balken in der Mitte).

Zu beachten gilt es dabei, dass die Modelle einen recht hohen Platzbedarf haben. Ein Modell «verschlingt» irgendwo zwischen 10 und 100 GByte, i.d.R. sind es etwa 50 GByte. Dafür sind aber die Resultate (z.B. mit Flux-dev) schon «verdammt» gut. In diesem Sinne, enjoy!

Hinweis: Im Prinzip läuft ComfyUI auch ohne AMD-Grafikkarte, sofern die CPU

für die Verarbeitung verwendet wird und genügend RAM zur Verfügung steht. Die Antwortzeiten sind aber sehr sehr gemächlich. 5 Sekunden HD-Video-Rendering mit Audio-Sync auf der R9700 benötigen aktuell einige Minuten, ohne GPU wären es ein paar Stunden. Ferner kann es je nach RAM-Situation (Hauptspeicher) notwendig sein, eine grosse Swap-Partition zu erstellen. Dies wird mit root mit **cd /var/lib/vz;dd if=/dev/zero of=avswap bs=64M count=1024;mkswap avswap** erreicht. Danach den Rechner neu starten und mit 'swapon' überprüfen, ob/dass die 64 GB aktiviert sind.
Hinweis II: Wer an dieser Stelle mehr Informationen zu ArchivistaDMS erwartet hätte, dem sei gesagt, dieser Teil folgt in einem der nächsten Blogs, keine Bange!