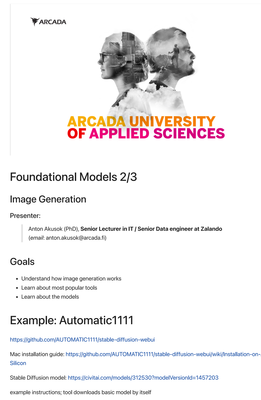


Foundational Models 2/3

Image Generation

Presenter:

> Anton Akusok (PhD), **Senior Lecturer in IT / Senior Data engineer at Zalando**
> (email: anton.akusok@arcada.fi)

Goals

- Understand how image generation works
- Learn about most popular tools
- Learn about the models

Example: Automatic1111

https://github.com/AUTOMATIC1111/stable-diffusion-webui

Mac installation guide: https://github.com/AUTOMATIC1111/stable-diffusion-webui/wiki/Installation-on-Apple-Silicon

Stable Diffusion model: https://civitai.com/models/312530?modelVersionId=1457203

example instructions; tool downloads basic model by itself

```
brew install cmake protobuf rust python@3.10 git wget
git clone https://github.com/AUTOMATIC1111/stable-diffusion-webui
```

download model into `stable-diffusion-webui/models/Stable-Diffusion`

```
cd stable-diffusion-webui
./webui.sh
```

Positive prompt:

- pet hamster dressed as an english gentleman in a fine suite, photorealistic, victorian style

Negative prompt:

- low resolution, painting, distorted

Comparing default model and improved model
https://civitai.com/models/312530/cyberrealistic-xl

Iterative image-to-image improvements

Read article at:

https://github.com/AUTOMATIC1111/stable-diffusion-webui/discussions/2473

Alternative: DiffusionBee demo

Very simple app with locally running models

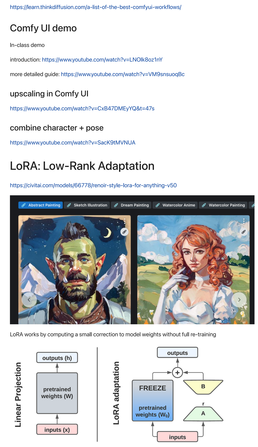
https://diffusionbee.com

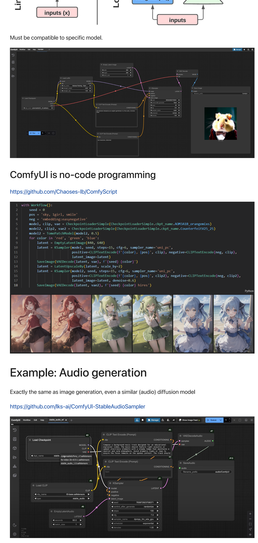
Example: Comfy UI ~ "AI Photoshop"

Advanced workflow system to generate any images one can imagine.

Combines multiple models and steps for detailed controls like posture.

Requires good model understanding; watch guides if you want to get serious in this field.





Installation:

- https://comfyui-wiki.com/en/install/install-comfyui/install-comfyui-on-mac

Models:

- General model download: https://comfyui-wiki.com/en/install/install-models/install-vae
- Model hosts: civitai.com, huggingface
- Example model checkpoint: https://civitai.com/models/443821/cyberrealistic-pony

Guides:

- Random guide https://stable-diffusion-art.com/comfyui/
- Workflow suggestions: https://www.reddit.com/r/StableDiffusion/comments/zch2kz/workflow_guide/
- Workflows available on Civitai.com together with models!

Comfy UI installation may complain about Mac access rights: go to Settings -> Privacy and Security -> enable access.

Comfy UI opens at an example workflow with comments how to use it.

- Read the instructions
- Download models (Stable Diffusion checkpoint and VAE)
- Place them in correct folders of Comfy UI created by the installer
- Restart Comfy UI and select your models

Set mask in image mask editor, then run the example workflow.

ComfyUI workflows

Kinda hard to explain, here is the best demo

https://learn.thinkdiffusion.com/a-list-of-the-best-comfyui-workflows/

Comfy UI demo

In-class demo

introduction: https://www.youtube.com/watch?v=LNOk8oz1nlr

more detailed guide: https://www.youtube.com/watch?v=VM9nrsuoqBc

upscaling in Comfy UI

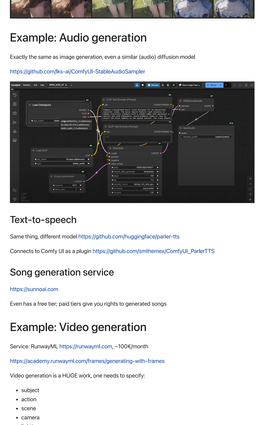
https://www.youtube.com/watch?v=CuB47DMEyYQ&t=47s

combine character + pose

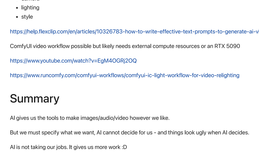
https://www.youtube.com/watch?v=SackX9MfvNUA

LoRA: Low-Rank Adaptation

https://civitai.com/models/66778/renoir-style-lora-for-anything-v50



LoRA works by computing a small correction to model weights without full re-training



Must be compatible to specific model.

ComfyUI is no-code programming

https://github.com/Chaoses-Ib/ComfyScript

Example: Audio generation

Exactly the same as image generation, even a similar (audio) diffusion model

https://github.com/fks-ajz/ComfyUI-StableAudioSampler

Text-to-speech

Same thing, different model https://github.com/huggingface/parler-tts

Connects to Comfy UI as a plugin https://github.com/smthemex/ComfyUI_ParlerTTS

Song generation service

https://sunnoai.com

Even has a free tier; paid tiers give you rights to generated songs

Example: Video generation

Service: RunwayML https://runwayml.com, ~10€/month

https://academy.runwayml.com/frames/generating-with-frames

Video generation is a HUGE work, one needs to specify:

- subject
- action
- scene
- camera
- lighting
- style

https://help.flexclip.com/en/articles/10326783-how-to-write-effective-text-prompts-to-generate-ai-videos

ComfyUI video workflow possible but likely needs external compute resources or an RTX 5090

https://www.youtube.com/watch?v=EgM4OGR2I0Q

https://www.runcomfy.com/comfyui-workflows/comfyui-ic-light-workflow-for-video-relighting

Summary

AI gives us the tools to make images/audio/video however we like.

But we must specify what we want, AI cannot decide for us - and things look ugly when AI decides.

AI is not taking our jobs. It gives us more work :D