



Module 8



Generative AI beyond Text



Generating Images



- Deconvolution and UNET model
- GANs
- AutoEncoders and VAEs
- Diffusion Models



Generating Art

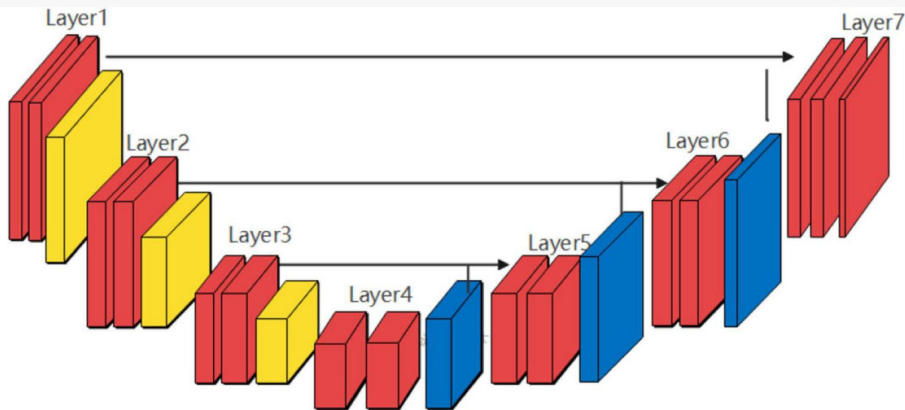


<https://beta.dreamstudio.ai/dream>


<https://lexica.art/aperture>



UNet

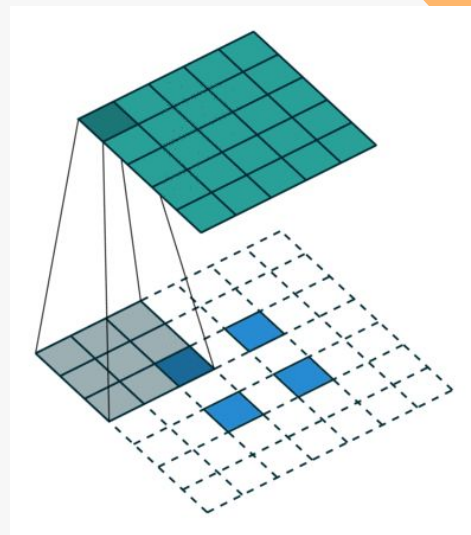


 Conv+BatchNormalization+ReLU

 Pooling operation

 Upsampling Layer

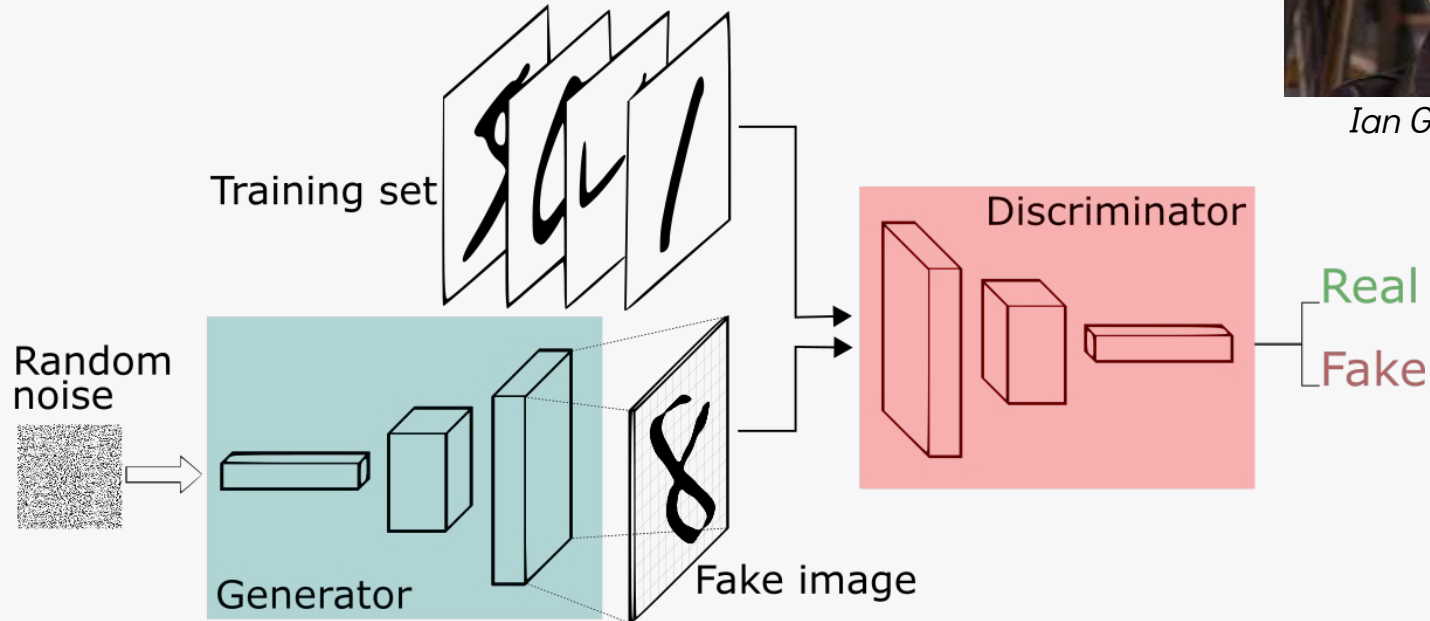
 Skip-Connection



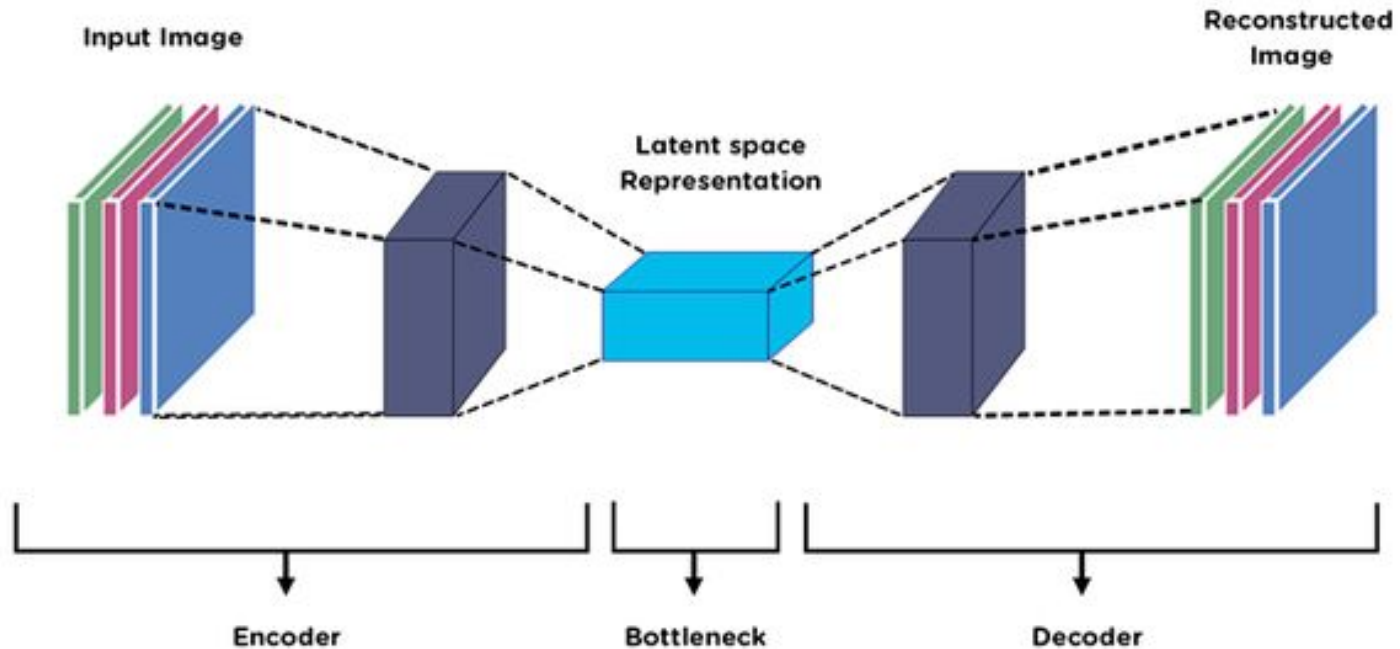
Generative Adversarial Networks



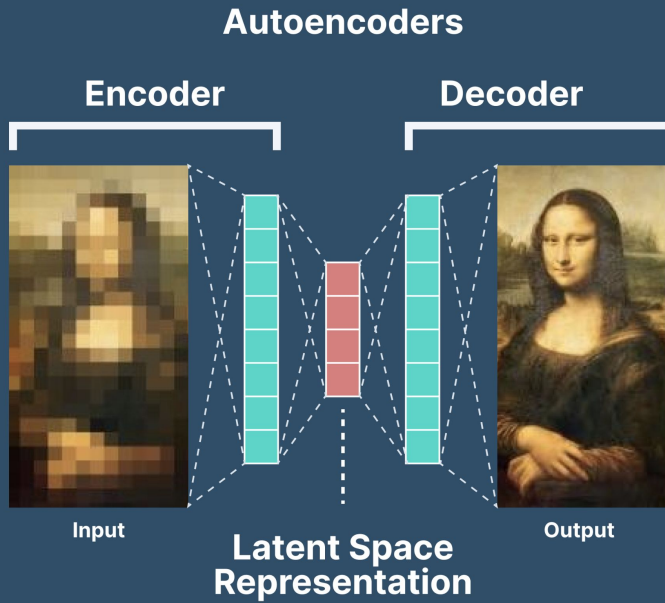
Ian GoodFellow



AutoEncoders

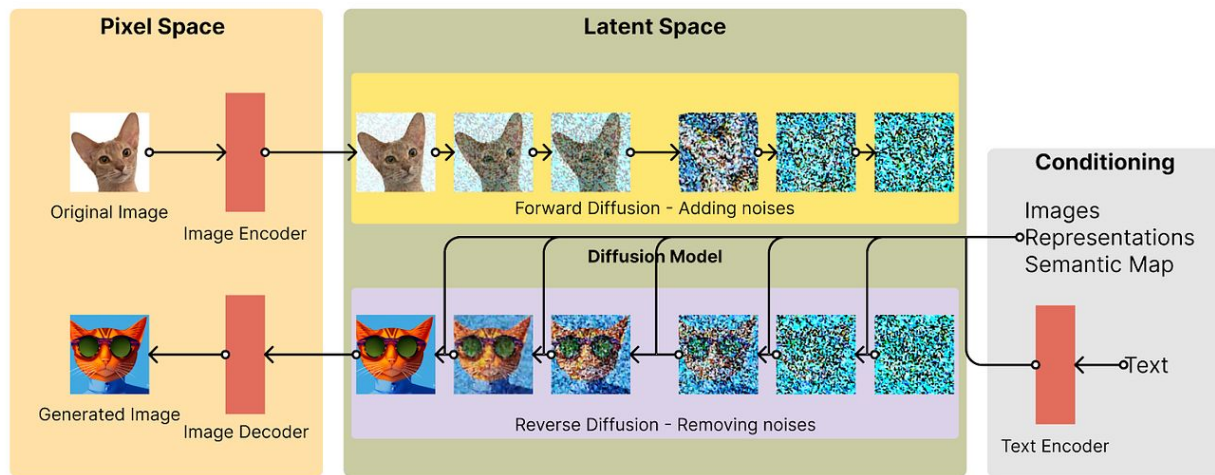


DeNoising AutoEncoder





Stable diffusion



[Stable Diffusion blogpost](#)

Future of AI



Beyond text-to-text and text-to-image!



Mitra at the Avatar X-Prize

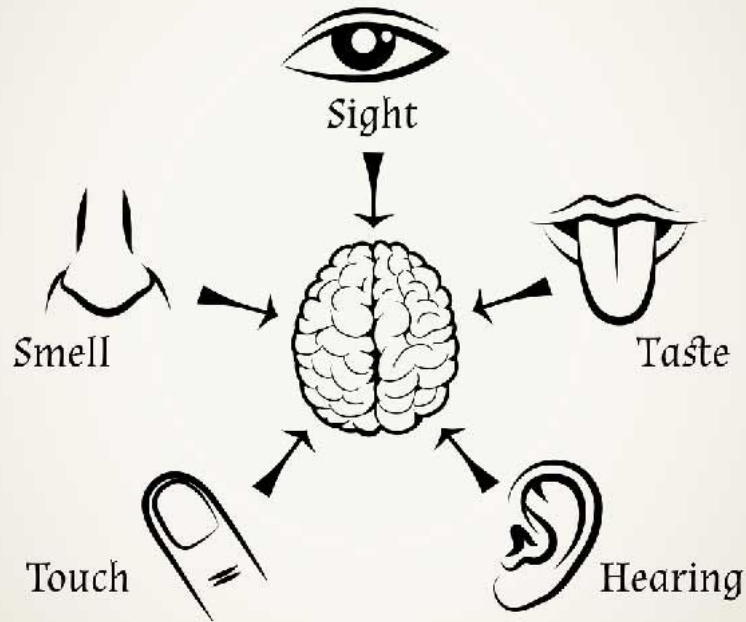


Can we bring multimodal interaction to people and robots: - Advanced touch - Tactile Sensing , Haptic Feedback



Multimodality - intro

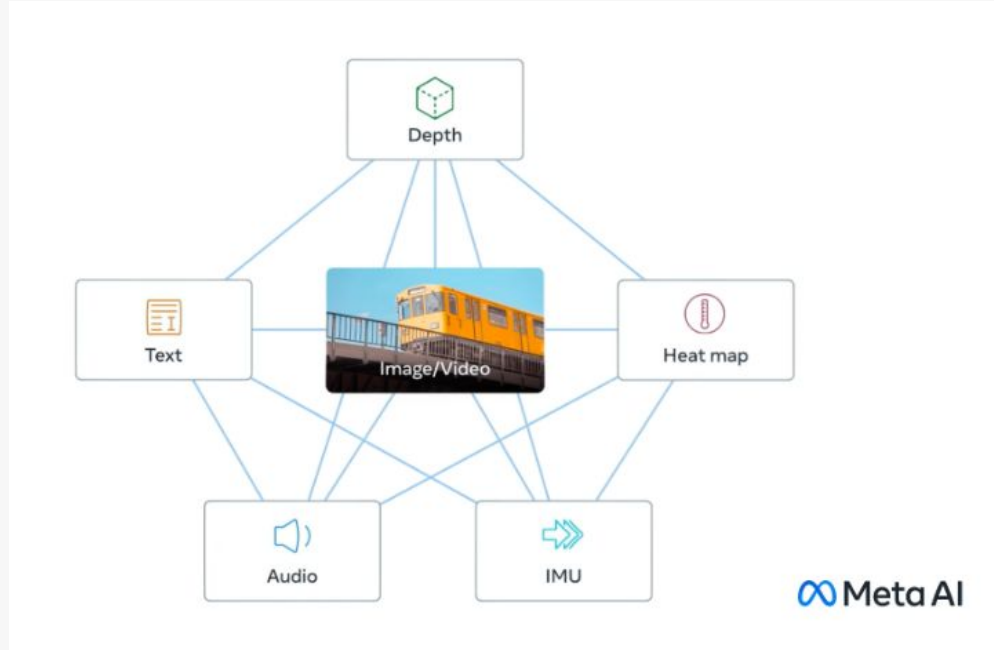
FIVE SENSES



*What other senses
can be digitized in
the future?*



Multimodality - today



<https://ai.facebook.com/blog/imagebind-six-modalities-binding-ai/>

MitraAI - Language meets Vision



- Image APIs for web and app developers
- Video Indexing and Search
- CCTV Applications



MitraRobot



- Elderly Care - Companionship and Monitoring
- Speech Recognition
- Speech Synthesis
- LLM based conversation engine
- Gesture recognition using body pose tracking
- Fall detection for safety monitoring
- Autonomous navigation and safety patrols



AI Hardware



- AWS EC2
- GOOGLE CLOUD
- MICROSOFT AZURE
- JARVIS LABS



The GPU revolution of AI



- NVIDIA and CUDA
- TSMC
- ASML
- INTEL & AMD
- ARM PROCESSORS
- APPLE M-Series CHIPS

