The Devil in the detAIl

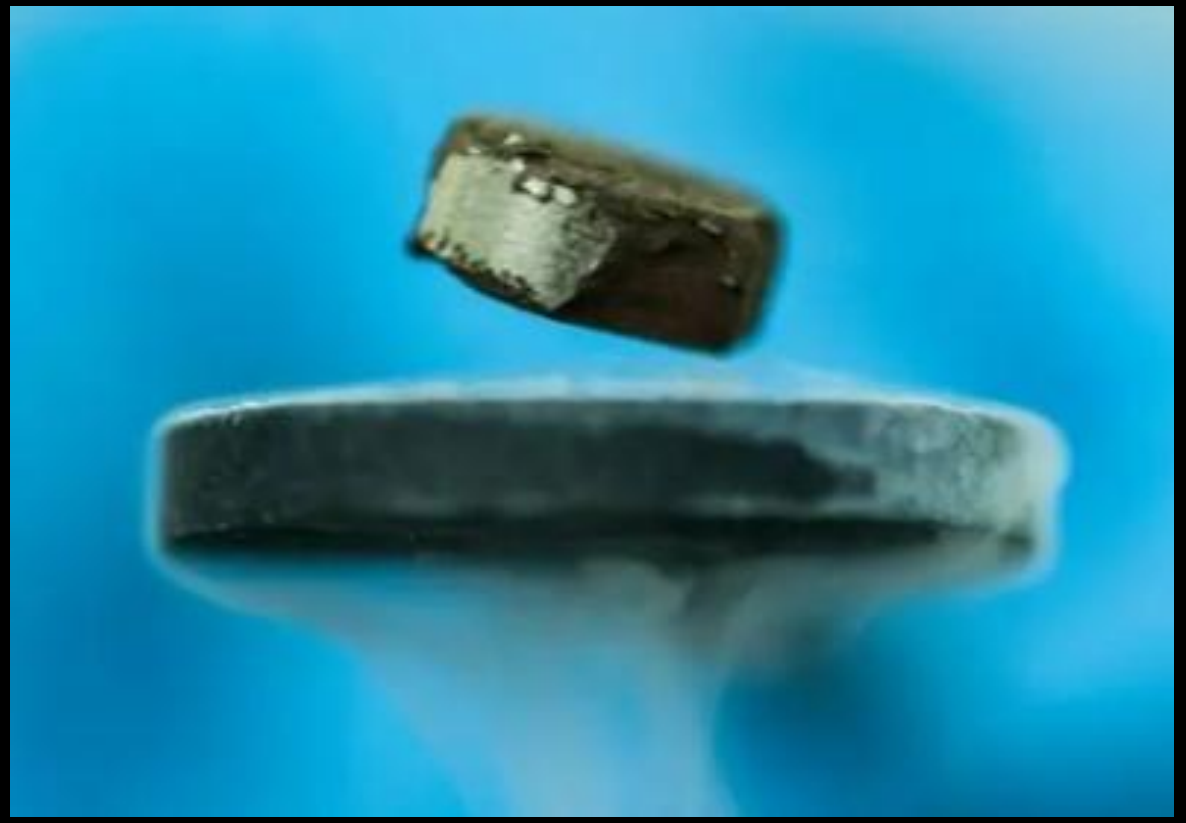
Breaking Through GenAI's Quality Ceiling

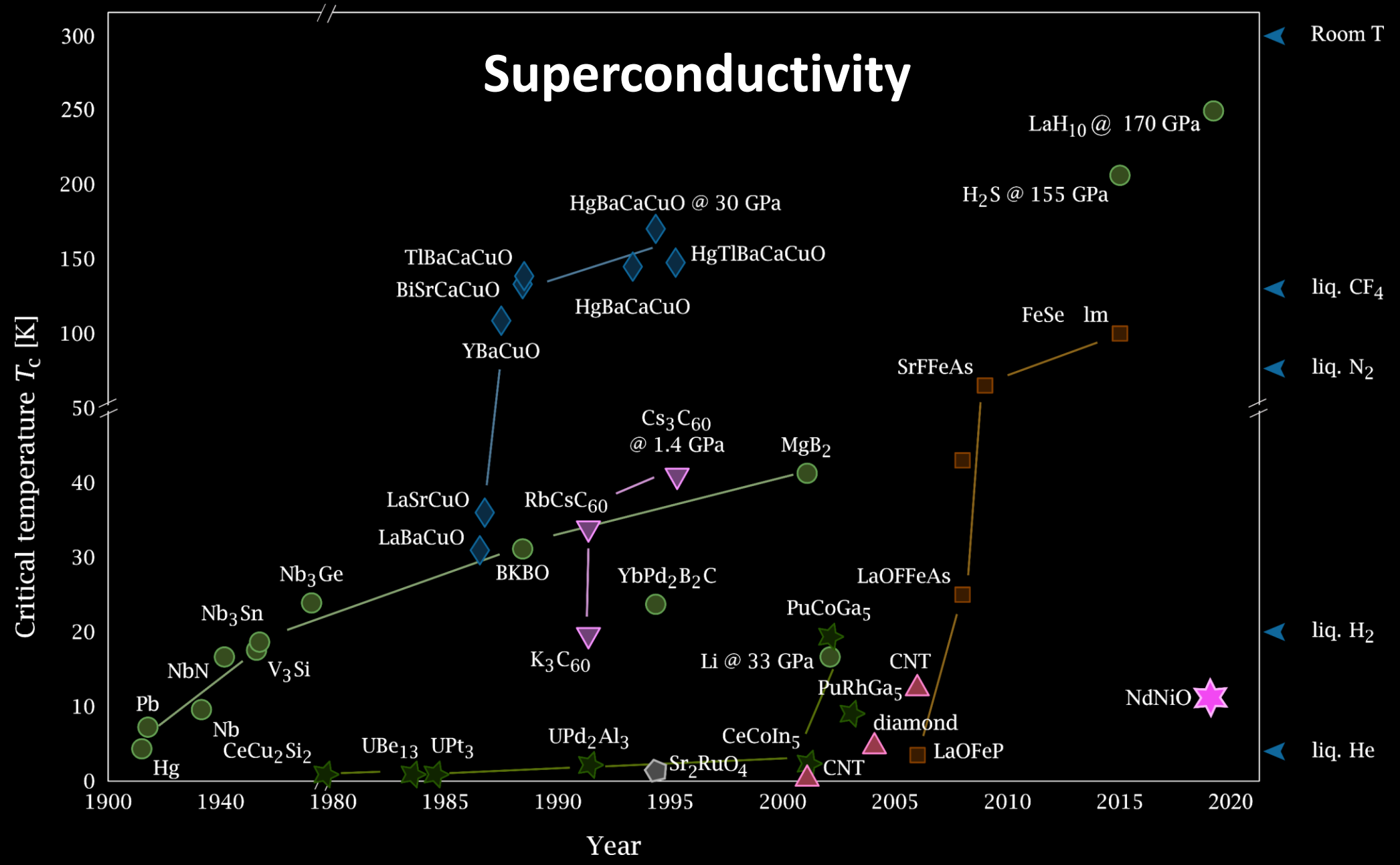
Daniel Sýkora

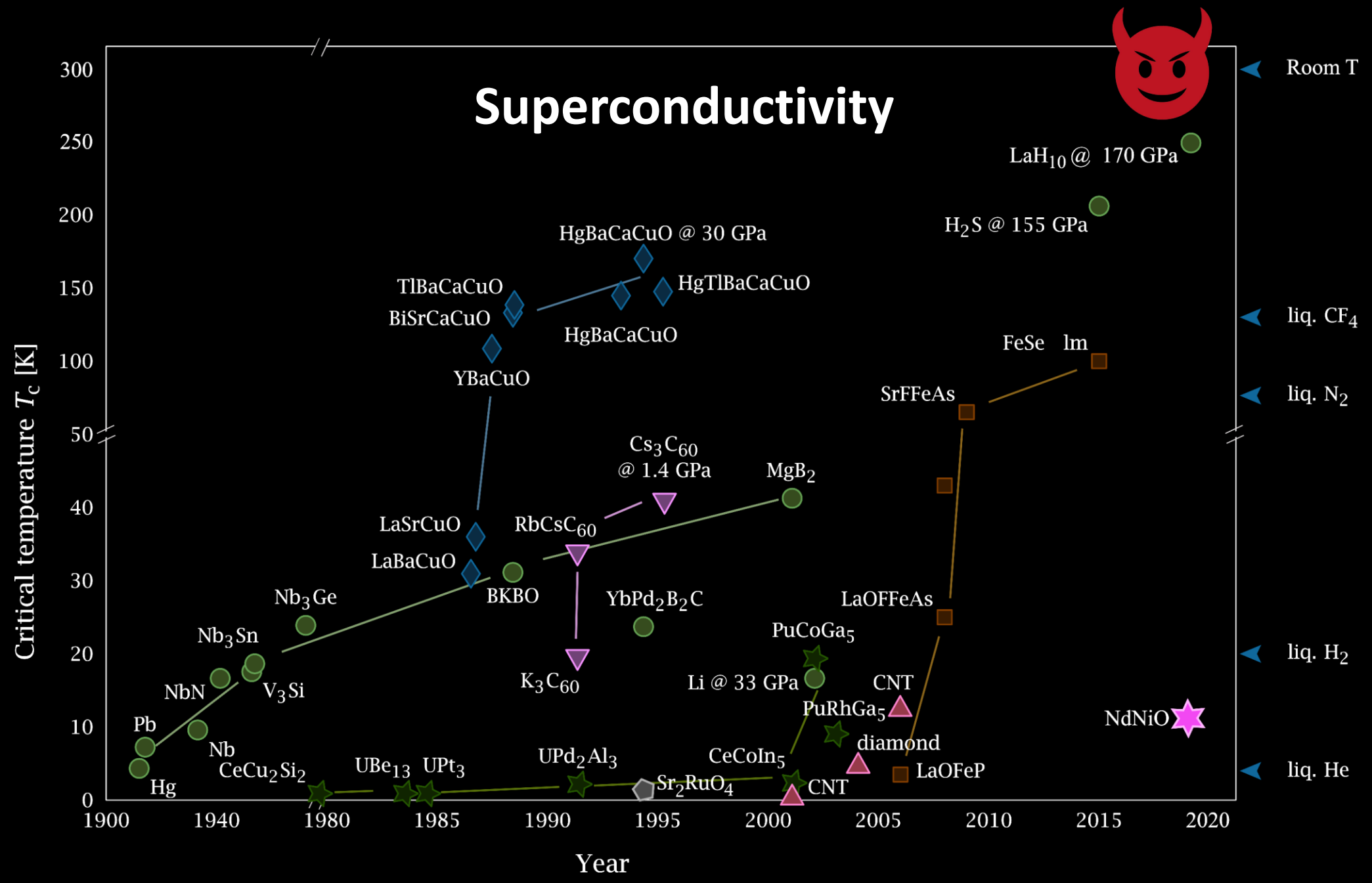
CTU
CZECH TECHNICAL UNIVERSITY IN PRAGUE

DCGI

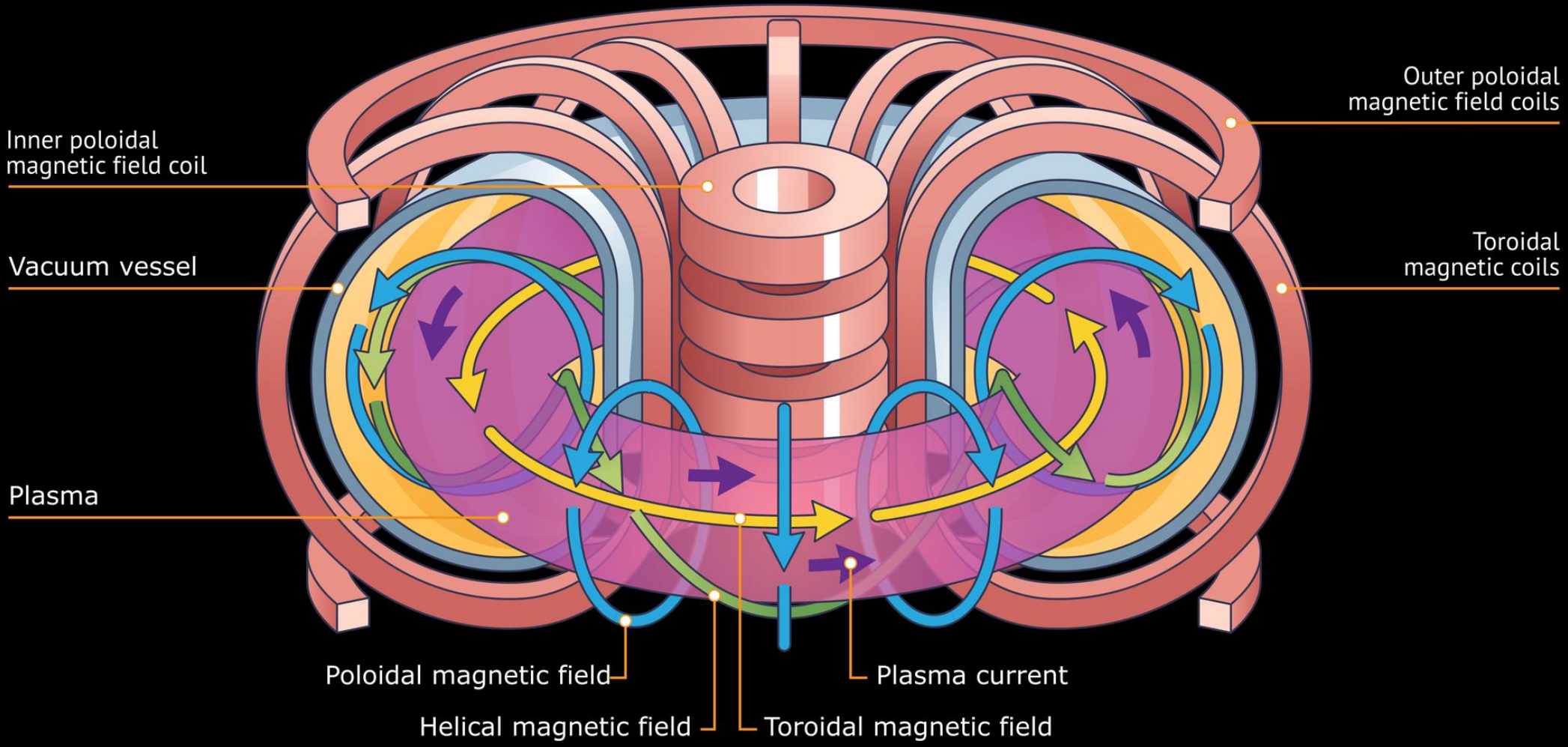
Superconductivity



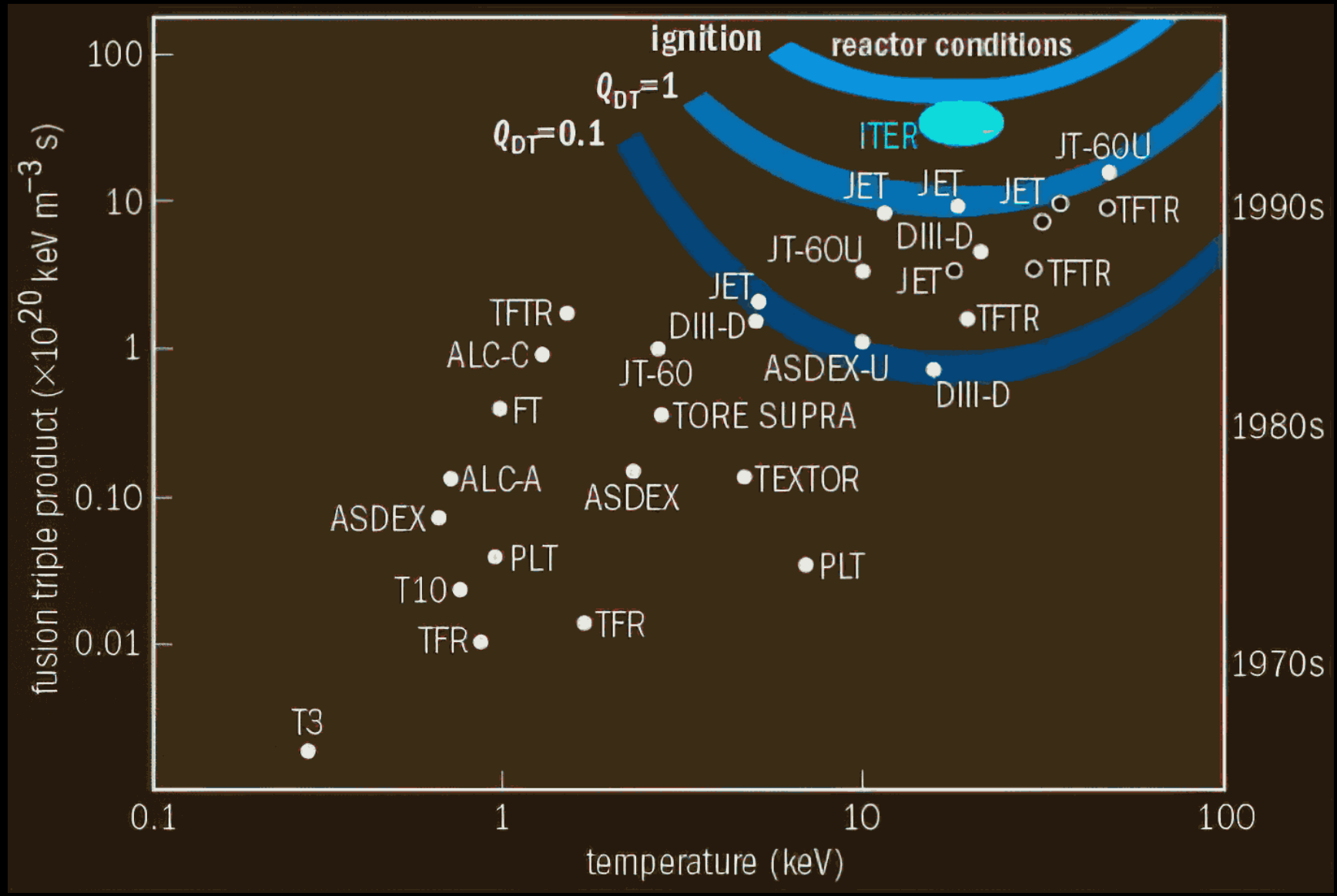

Superconductivity
Critical temperature T_c [K]
Year
300
250
200
150
100
50
40
30
20
10
0
1900
1940
1980
1985
1990
1995
2000
2005
2010
2015
2020
Room T
liq. CF_4
liq. N_2
liq. H_2
liq. He
LaH_{10} @ 170 GPa
H_2S @ 155 GPa
HgBaCaCuO @ 30 GPa
TlBaCaCuO
HgTlBaCaCuO
BiSrCaCuO
HgBaCaCuO
YBaCuO
FeSe lm
SrFFeAs
Cs_3C_{60} @ 1.4 GPa
MgB_2
LaSrCuO
$RbCsC_{60}$
LaBaCuO
BKBO
Nb_3Ge
$YbPd_2B_2C$
LaOFFeAs
Nb_3Sn
$PuCoGa_5$
K_3C_{60}
Li @ 33 GPa
CNT
NbN
V_3Si
$PuRhGa_5$
NdNiO
Pb
Nb
diamond
$CeCu_2Si_2$
UBe_{13}
UPt_3
UPd_2Al_3
$CeCoIn_5$
LaOFeP
Hg
Sr_2RuO_4
CNT


Superconductivity
Critical temperature T_c [K]
Year
300
250
200
150
100
50
40
30
20
10
0
1900
1940
1980
1985
1990
1995
2000
2005
2010
2015
2020
Room T
liq. CF_4
liq. N_2
liq. H_2
liq. He
LaH_{10} @ 170 GPa
H_2S @ 155 GPa
HgBaCaCuO @ 30 GPa
TlBaCaCuO
HgTlBaCaCuO
BiSrCaCuO
HgBaCaCuO
YBaCuO
FeSe lm
SrFFeAs
Cs_3C_{60} @ 1.4 GPa
MgB_2
LaSrCuO
$RbCsC_{60}$
LaBaCuO
BKBO
Nb_3Ge
$YbPd_2B_2C$
LaOFFeAs
Nb_3Sn
$PuCoGa_5$
K_3C_{60}
Li @ 33 GPa
CNT
NbN
V_3Si
$PuRhGa_5$
NdNiO
Pb
Nb
diamond
$CeCu_2Si_2$
UBe_{13}
UPt_3
UPd_2Al_3
$CeCoIn_5$
LaOFeP
Hg
Sr_2RuO_4
CNT

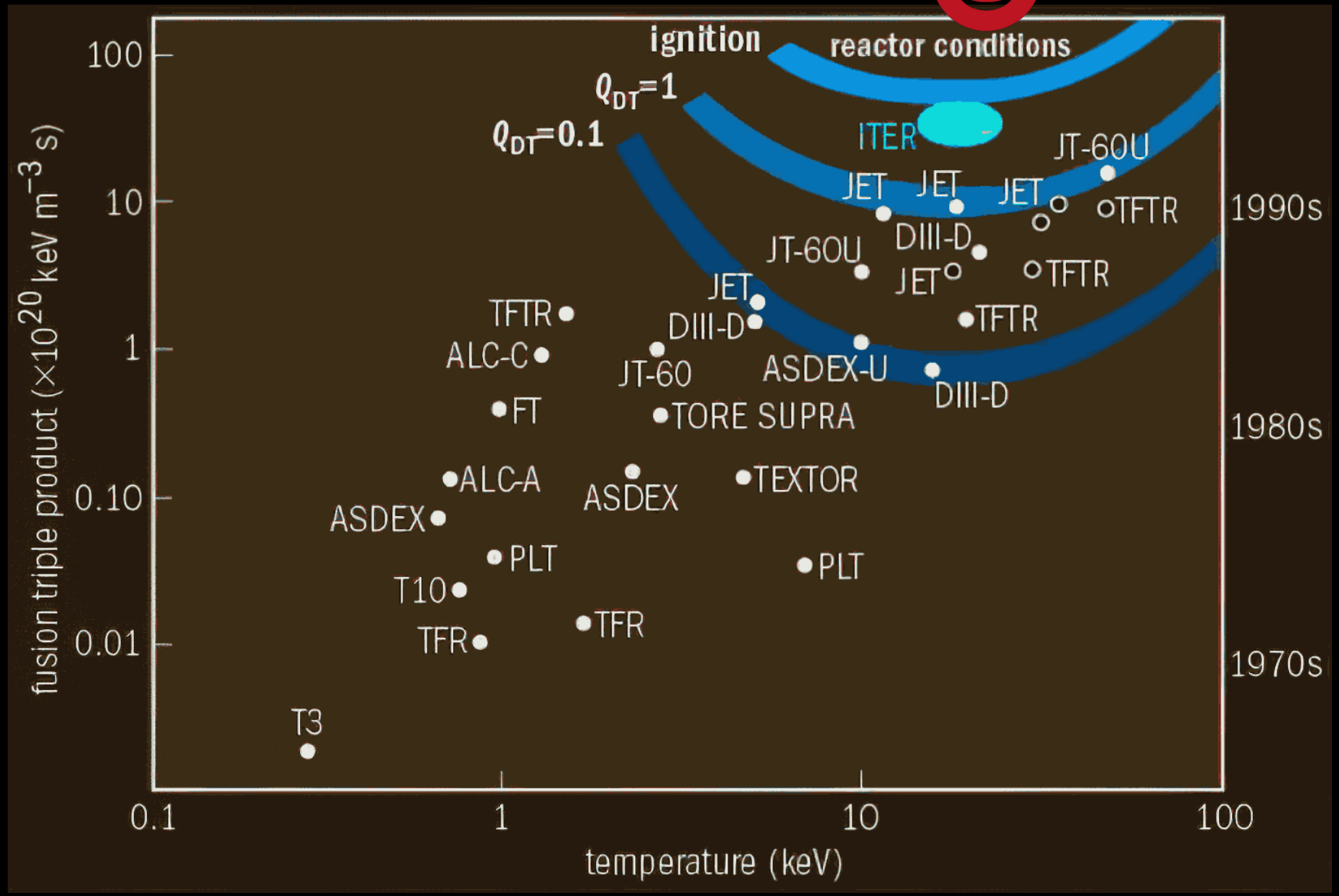
Nuclear Fusion



Nuclear Fusion



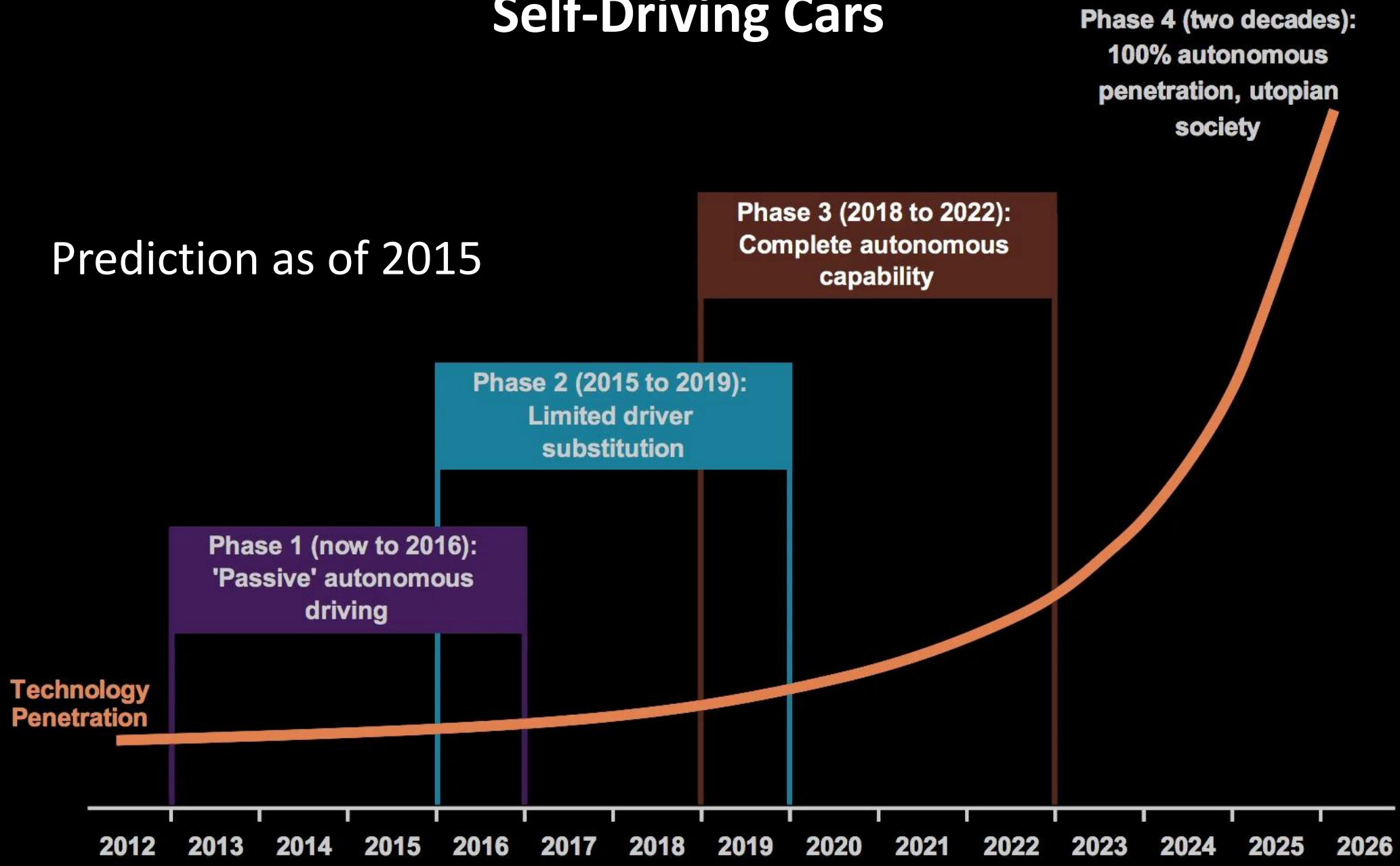
Nuclear Fusion



Self-Driving Cars

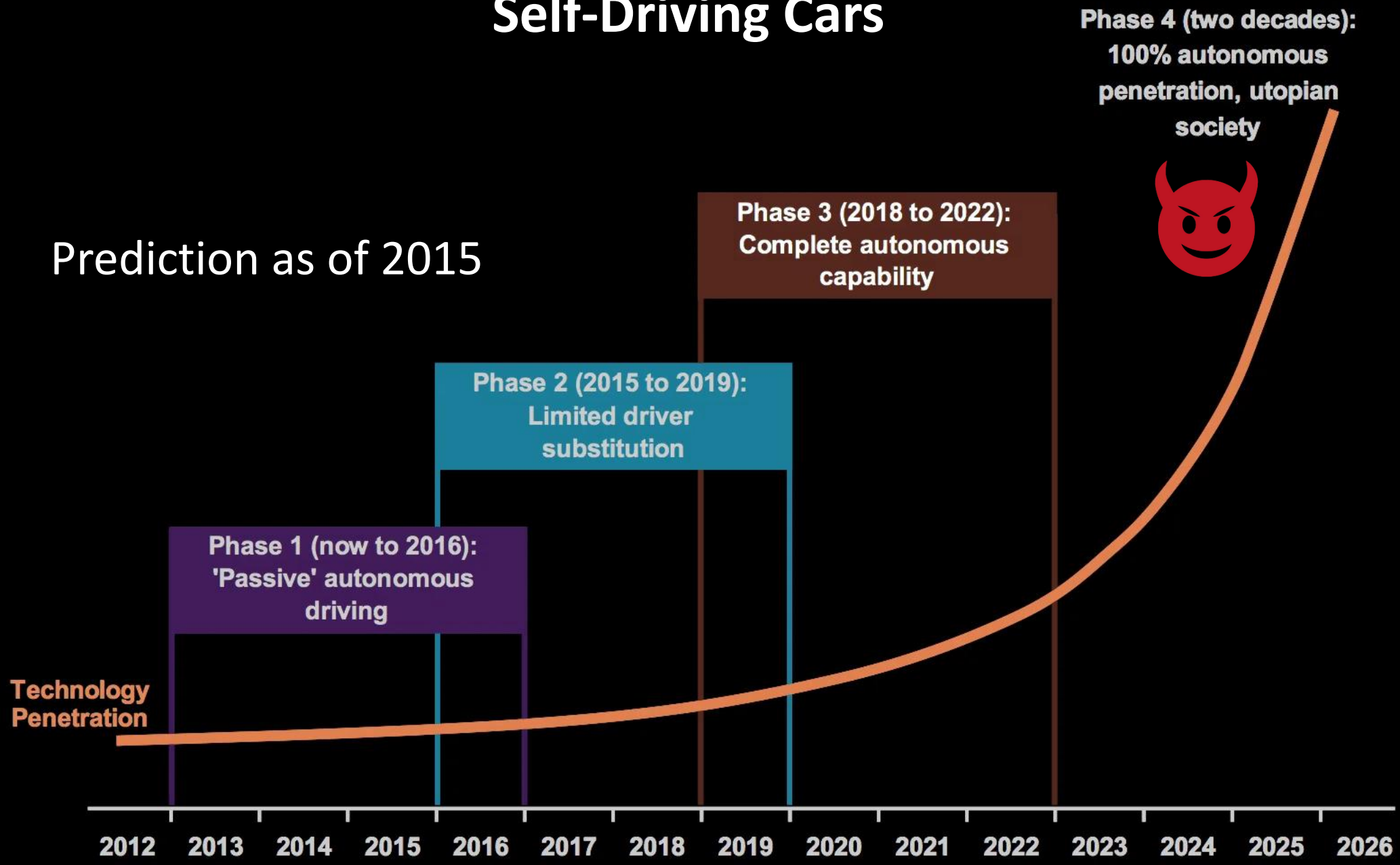
Self-Driving Cars

Prediction as of 2015

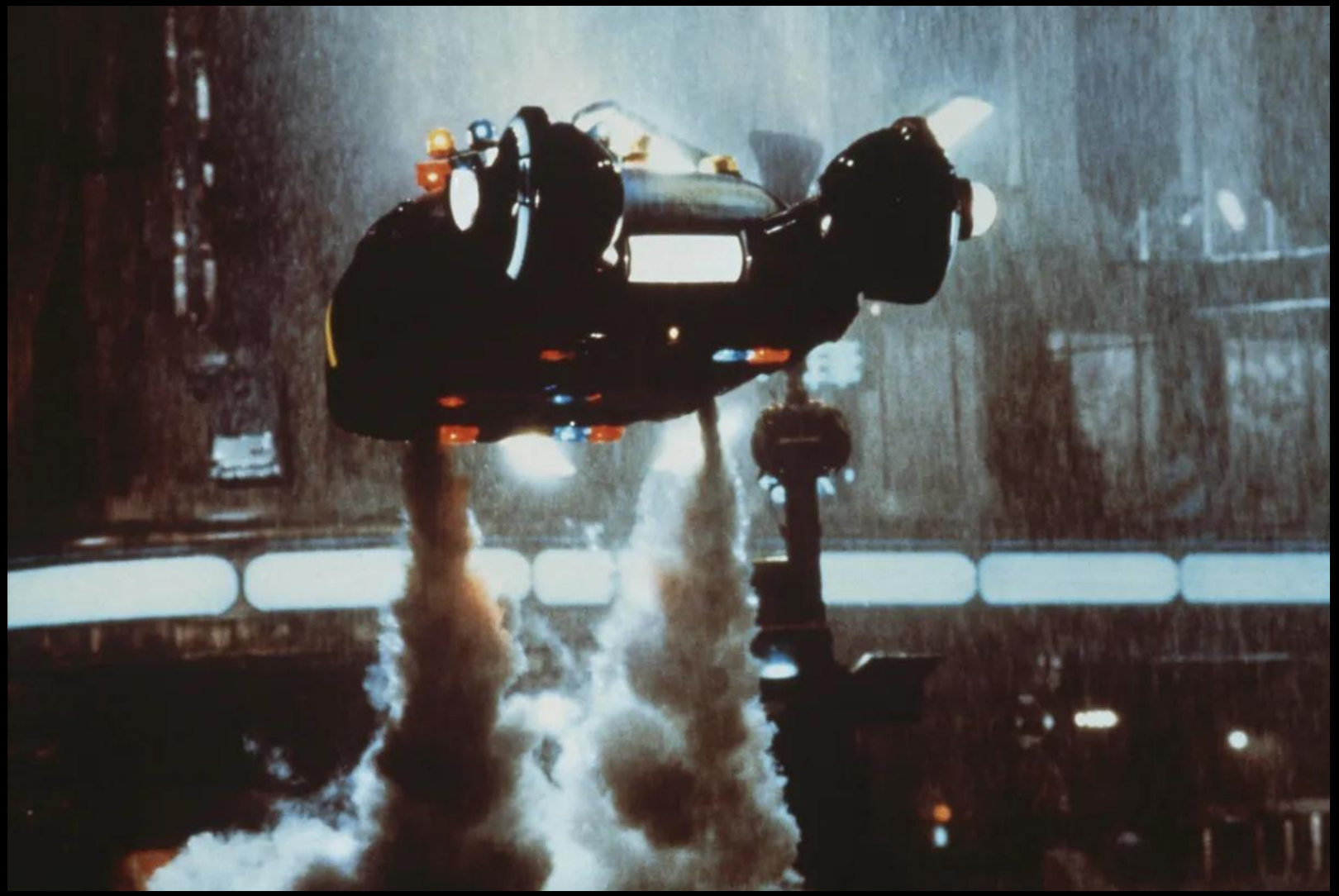


Self-Driving Cars

Prediction as of 2015



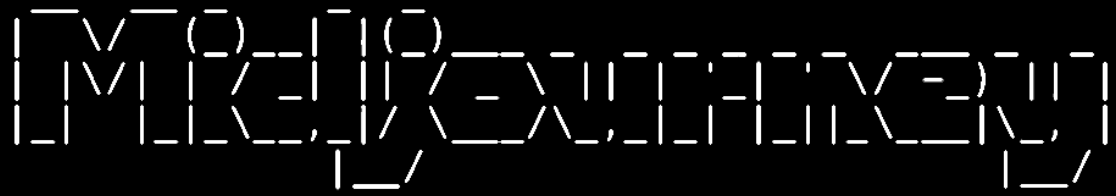
Blade Runner (2019)



Prediction as of 1982

Text-Based Image Generation

DALL·E



Stable Diffusion

…

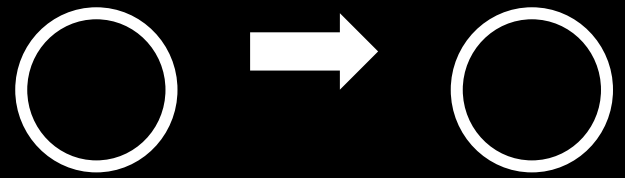
“a propaganda poster depicting a cat dressed as french emperor napoleon holding a piece of cheese”

“an espresso machine that makes coffee from human souls, artstation”

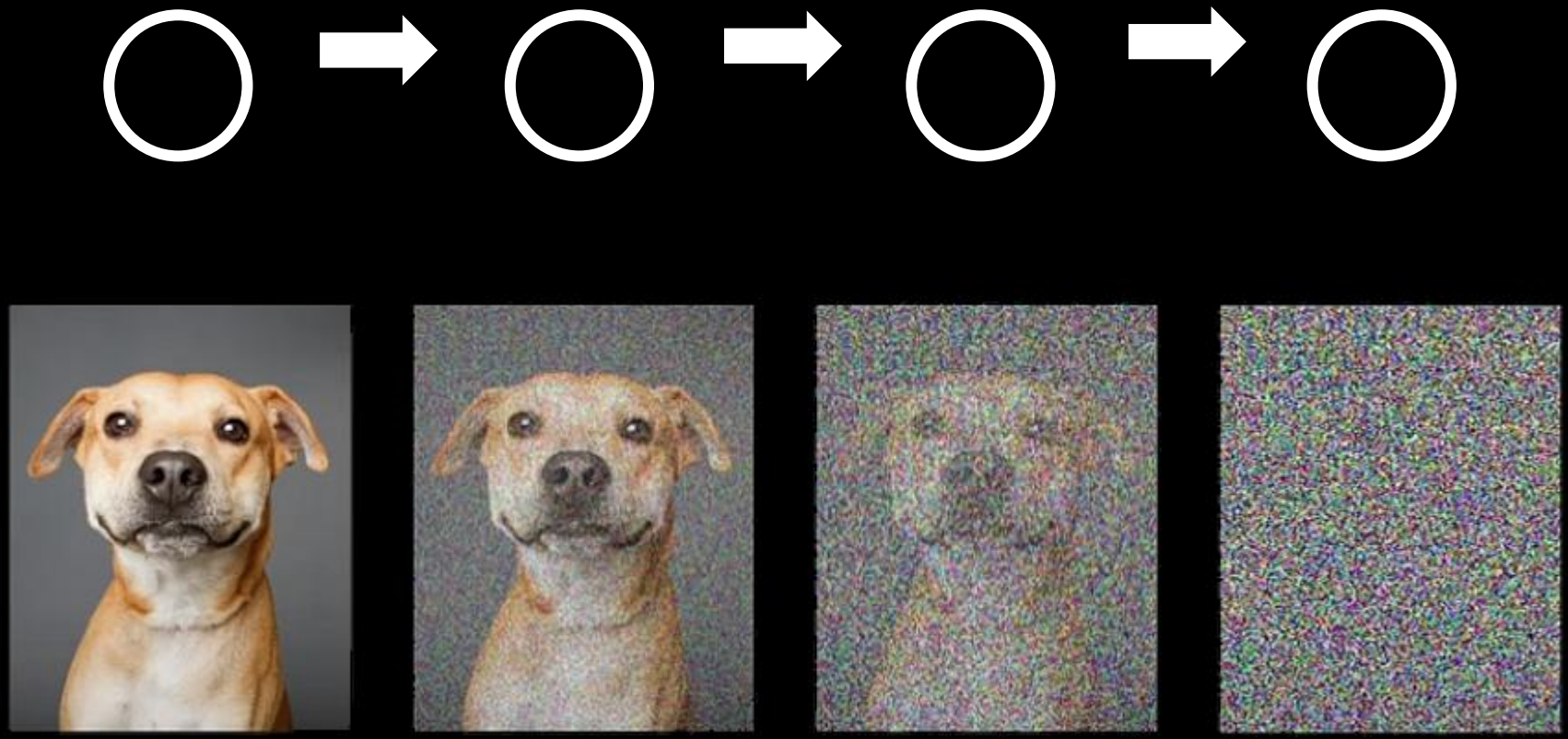
“teddy bears shopping groceries in the style of ukiyo-e"

“a bowl of soup that looks like a monster spray-painted on a wall”

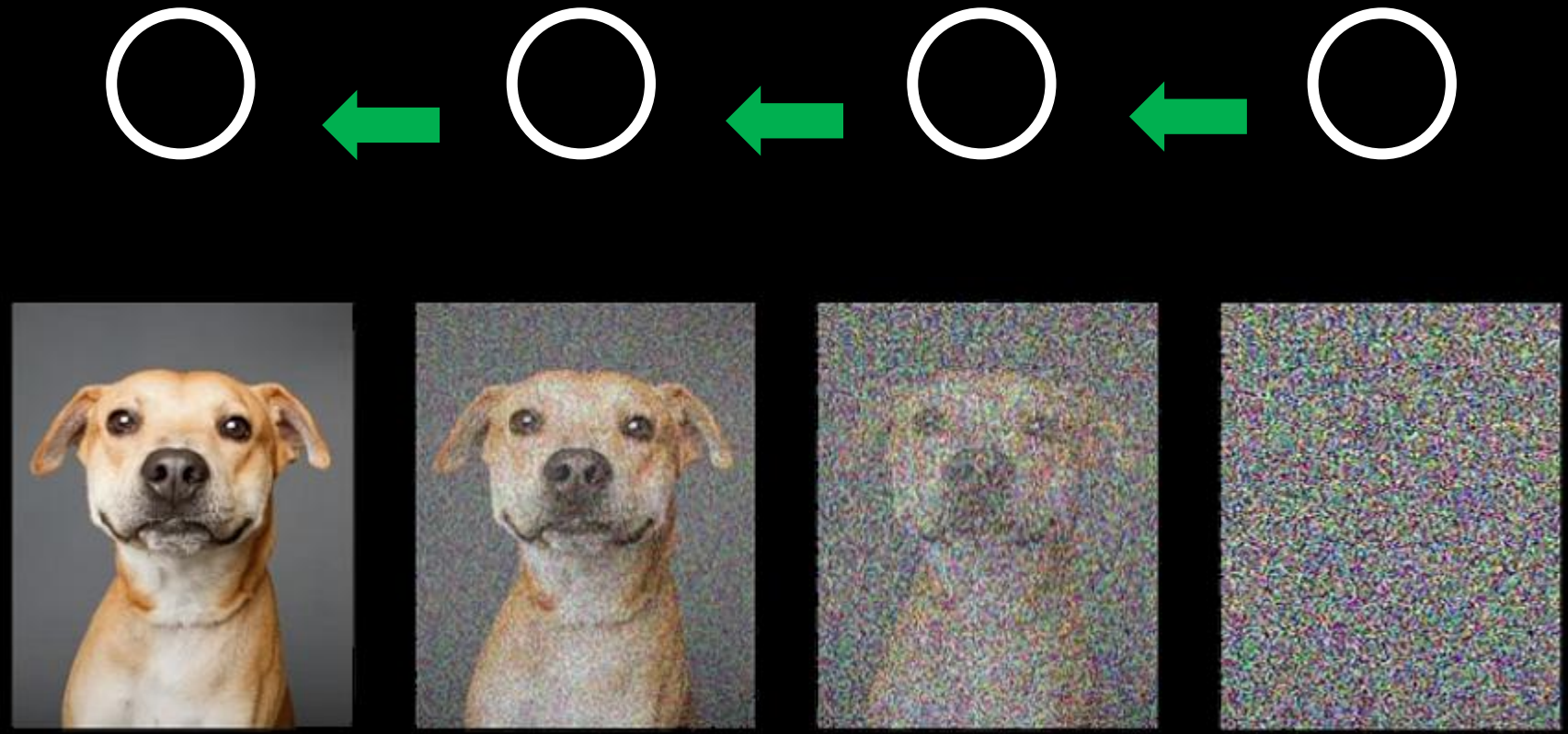
Diffusion Model



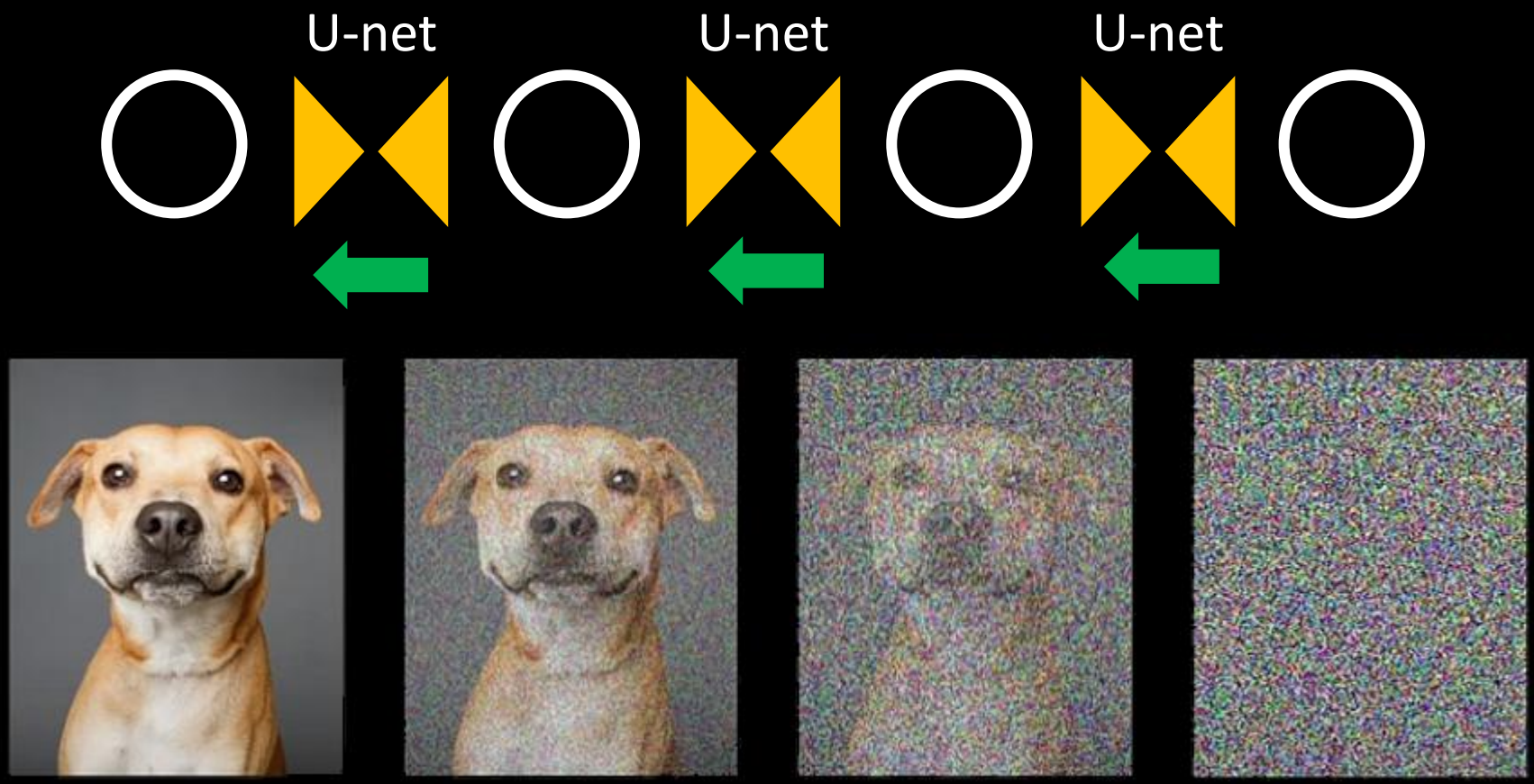
Diffusion Model



Diffusion Model



Diffusion Model



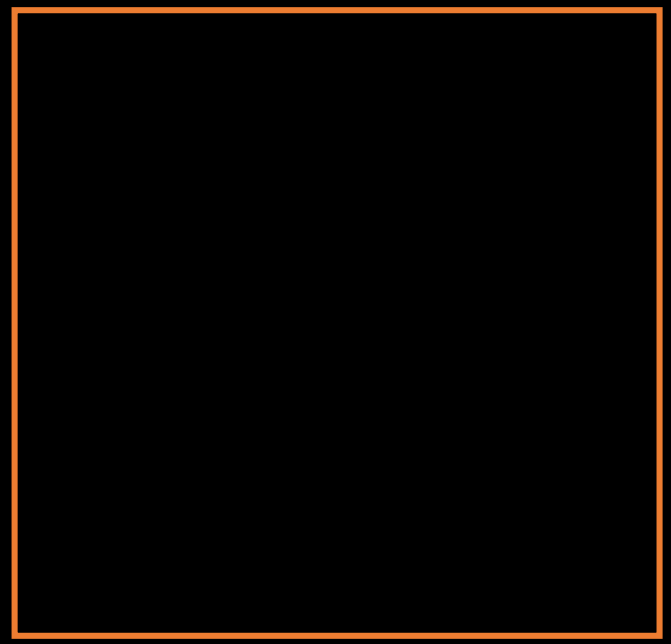
Diffusion Model

CLIP: Contrastive Language-Image Pre-training

= dog

CLIP: Contrastive Language-Image Pre-training

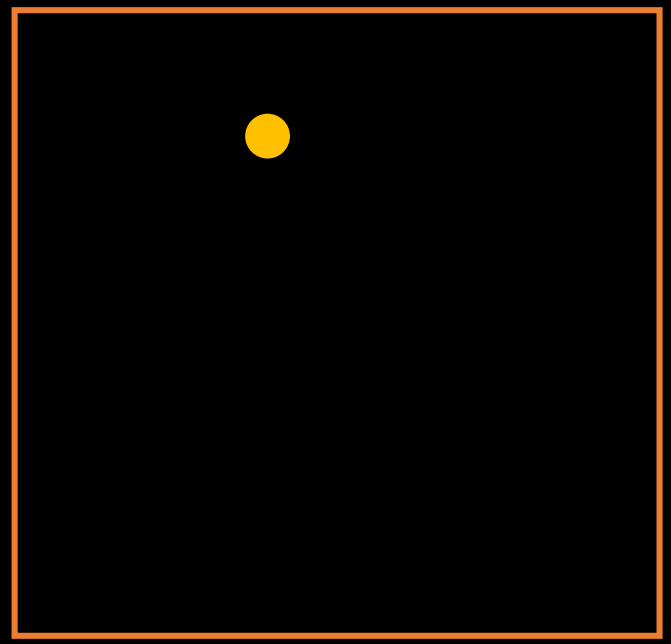
vector space



dog

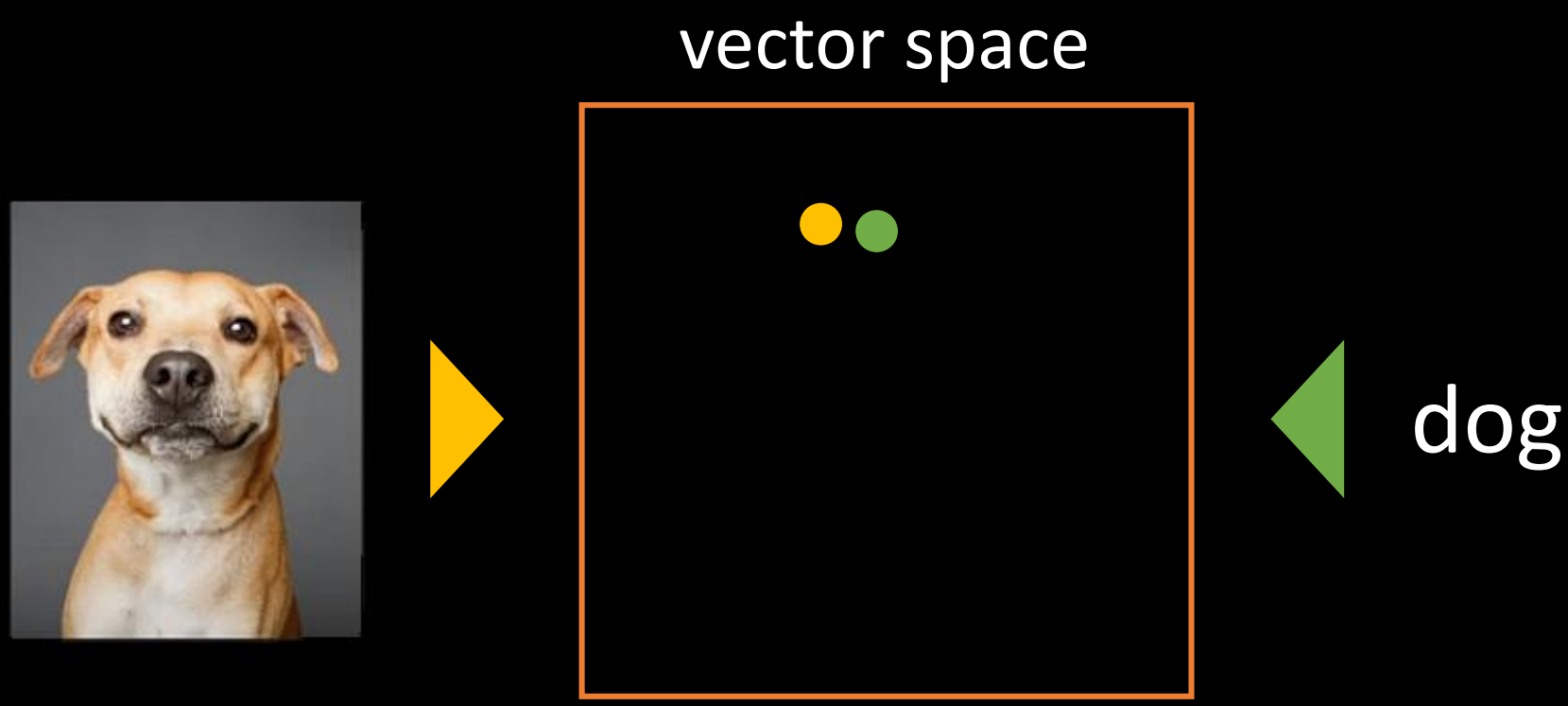
CLIP: Contrastive Language-Image Pre-training

vector space



dog

CLIP: Contrastive Language-Image Pre-training



CLIP: Contrastive Language-Image Pre-training

CLIP: Contrastive Language-Image Pre-training

 + ( - ) =

CLIP: Contrastive Language-Image Pre-training

 + ( - ) = 

CLIP: Contrastive Language-Image Pre-training

 + ( - ) = 

three cats + (dog – cat) = three dogs

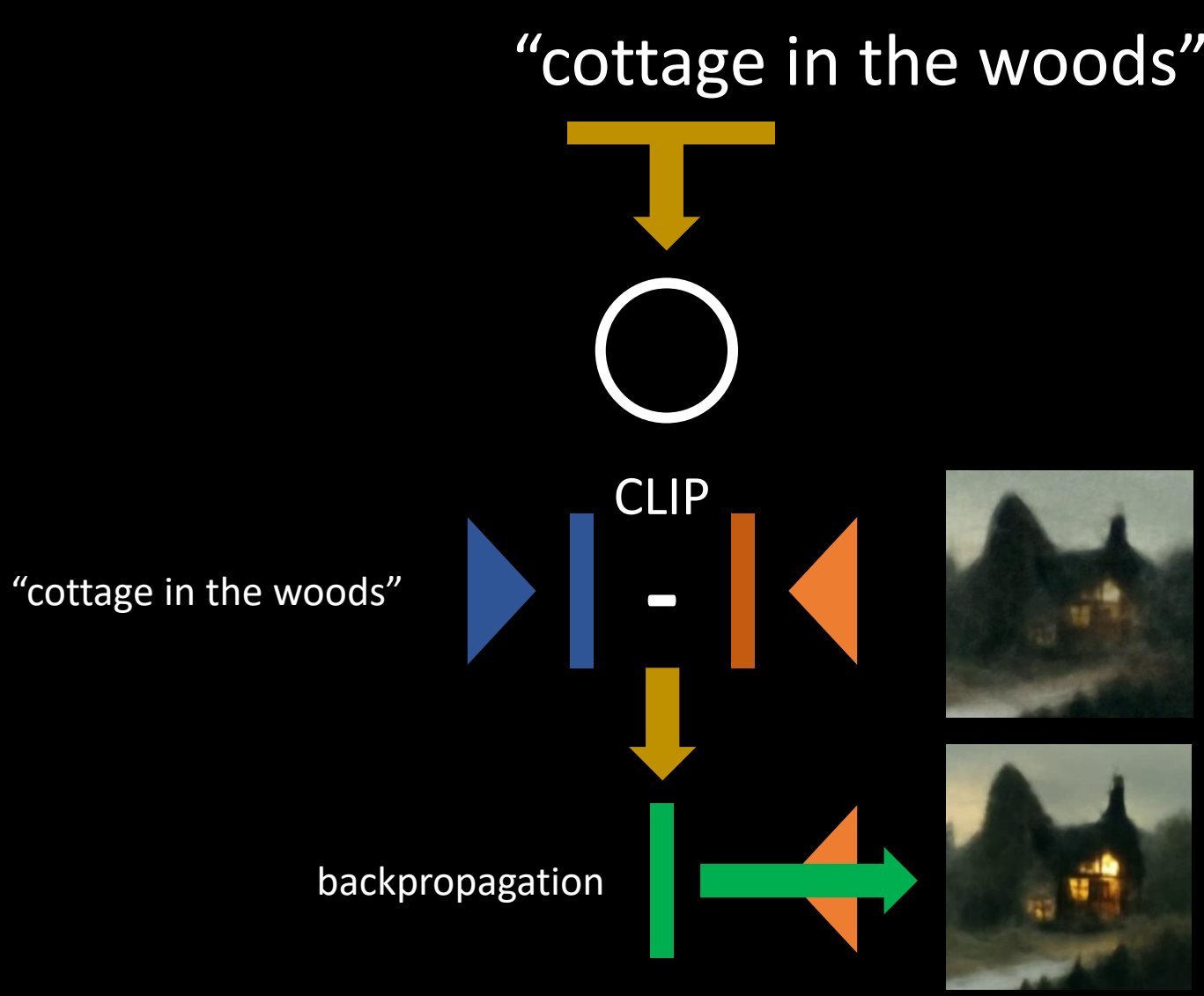
CLIP: Contrastive Language-Image Pre-training

 + ( - ) = 

three cats + (dog – cat) = three dogs

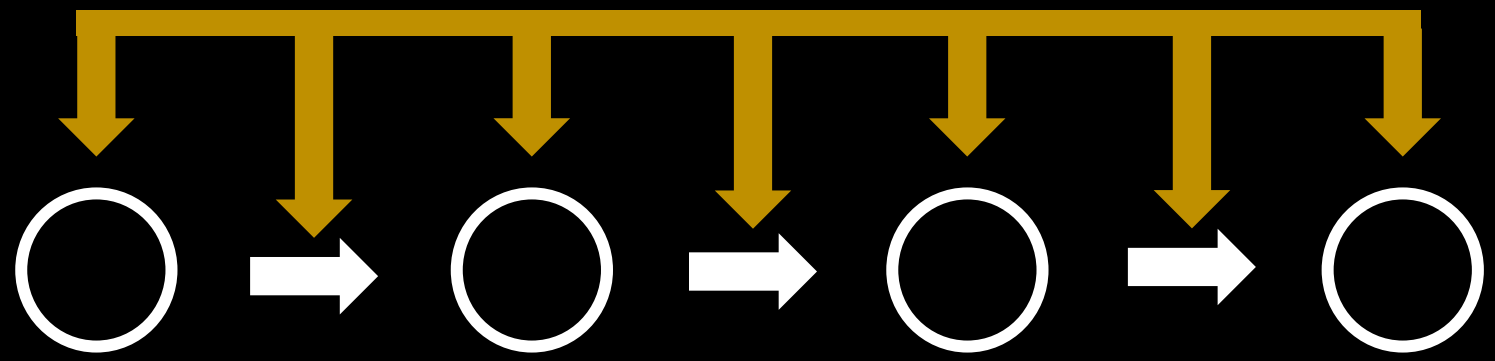
 + (dog – cat) = 

Text-Guided Diffusion Model



Text-Guided Diffusion Model

“cottage in the woods”



GLIDE: Text-Guided Diffusion Model (2021)

The Devil in the detAIl

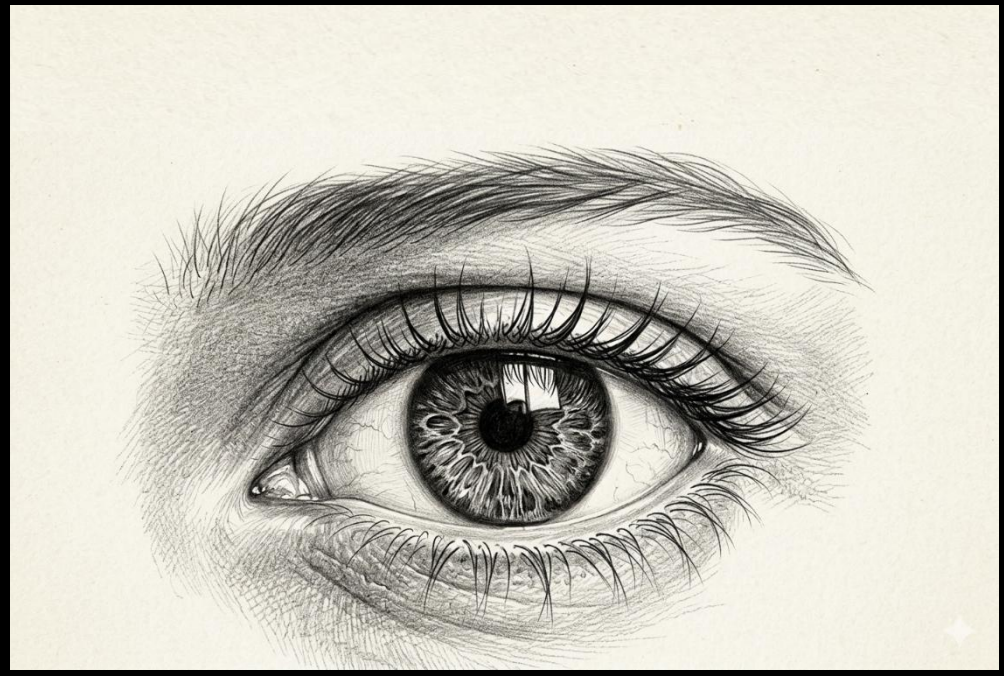
The Devil in the detAIl

Real Pencil Drawing

Generated using Midjourney

Real Pencil Drawing

Generated using Nano Banana 2

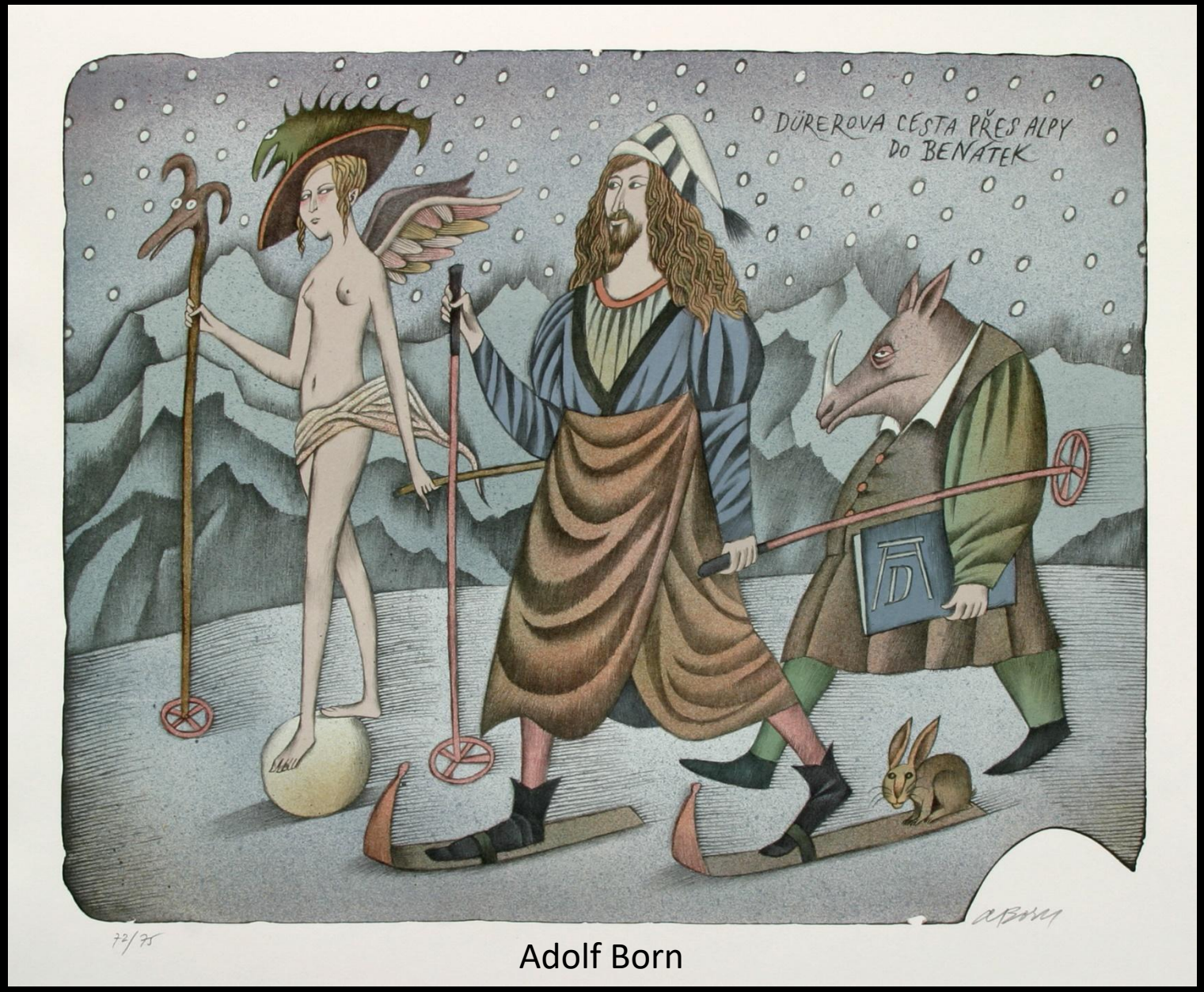


Concept Art

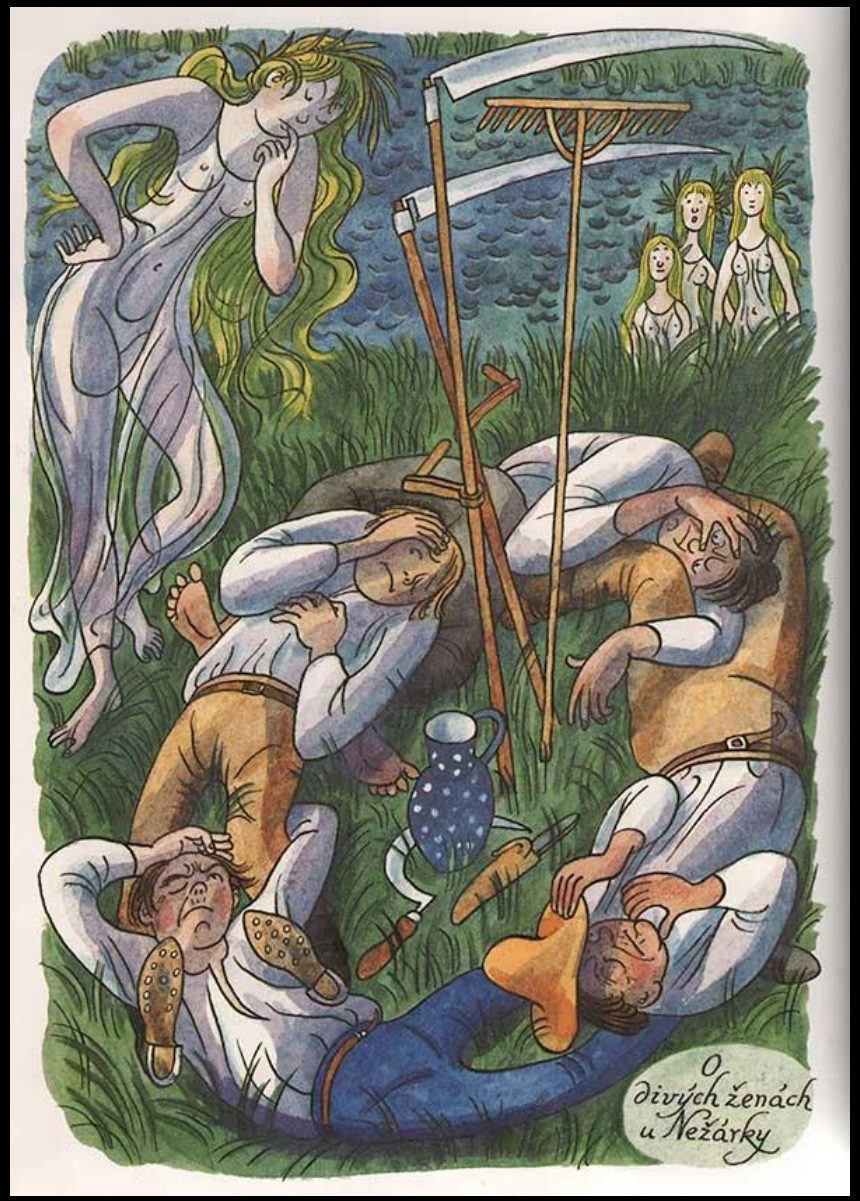
Final Product

Josef Lada

Josef Čapek

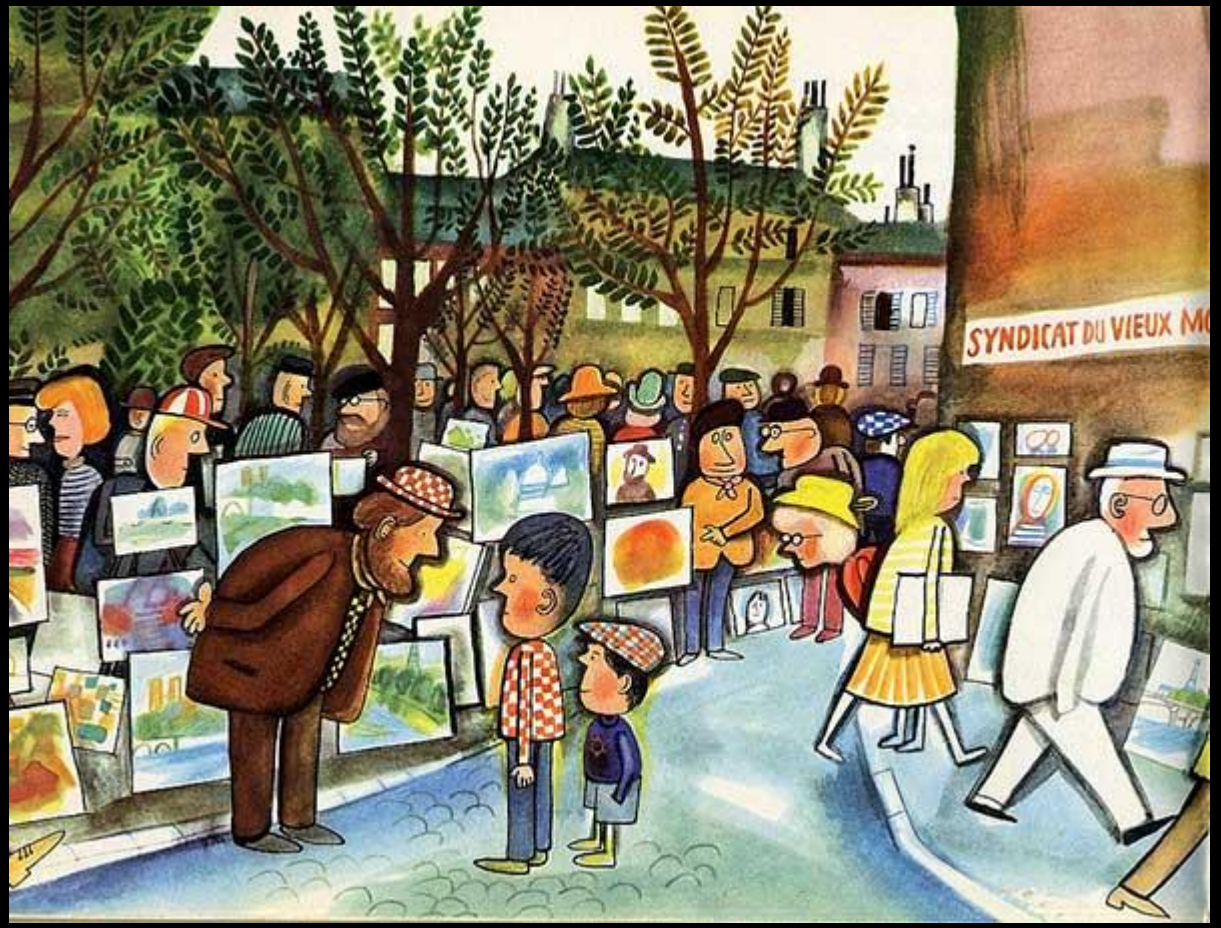


Adolf Born



Cyril Bouda

Ondřej Sekora



Jiří Kalousek

Jiří Šalamoun

Jiří Trnka

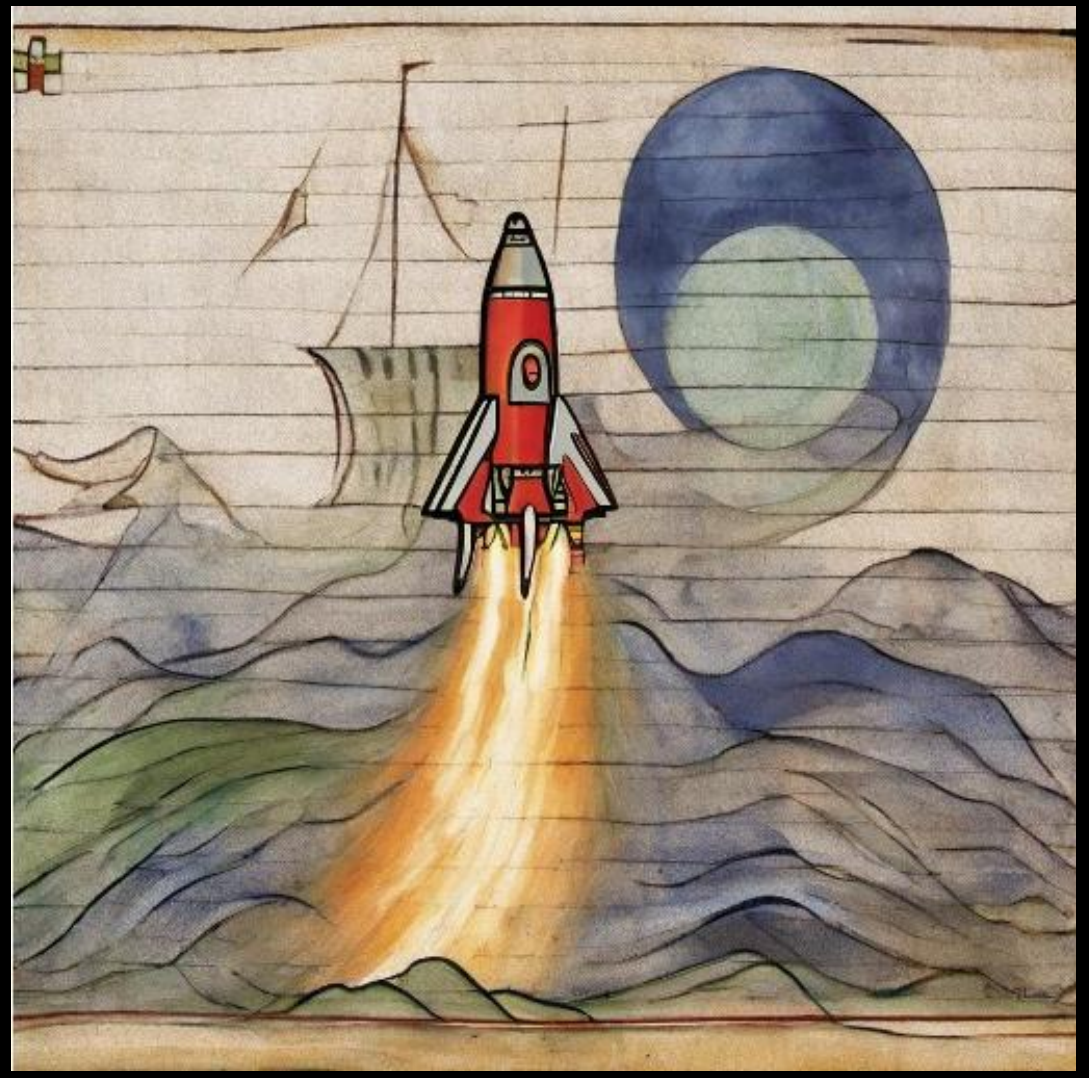
František Skála

Generative Fill (Adobe Firefly)

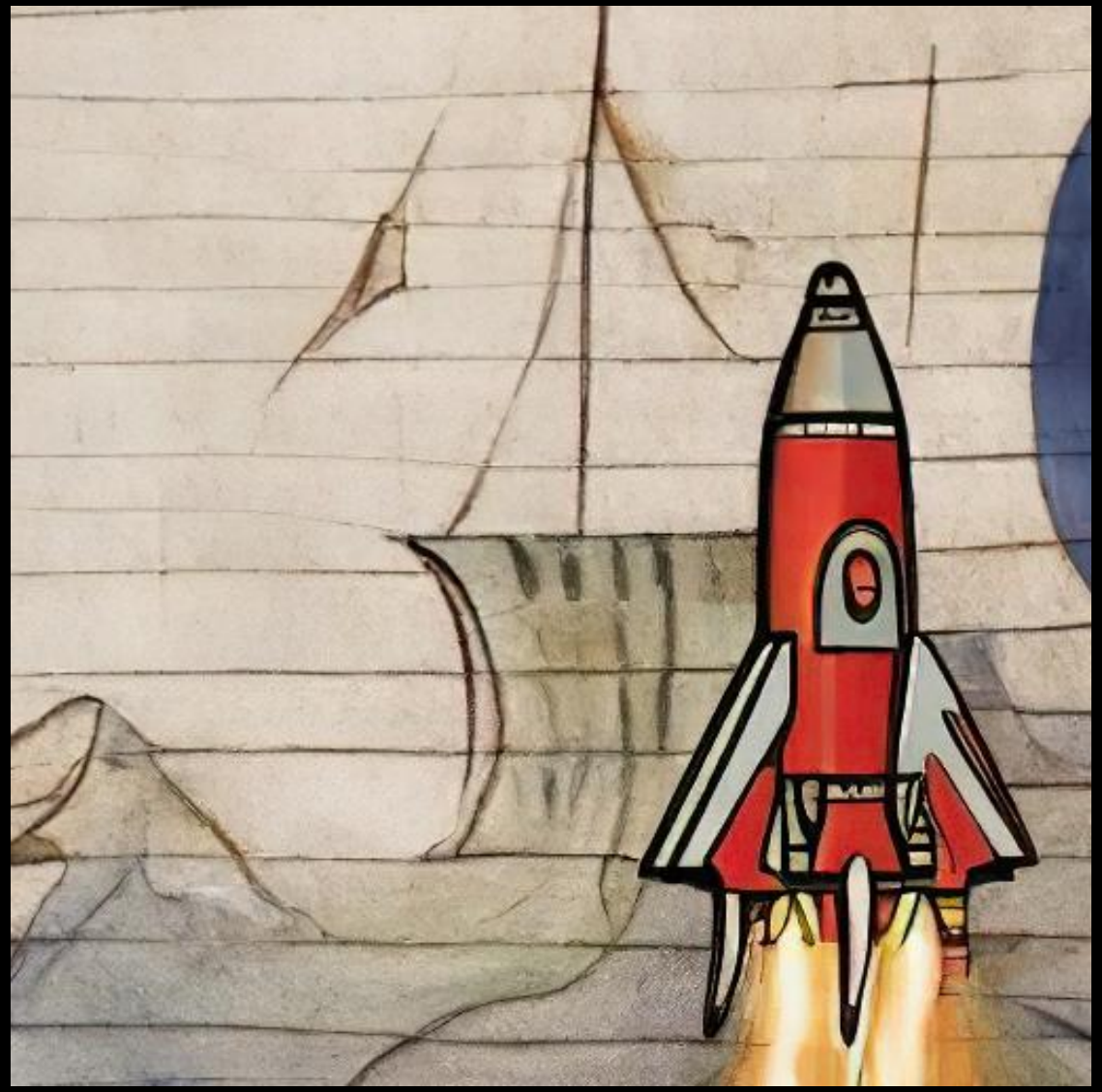
Generative Fill (Adobe Firefly)

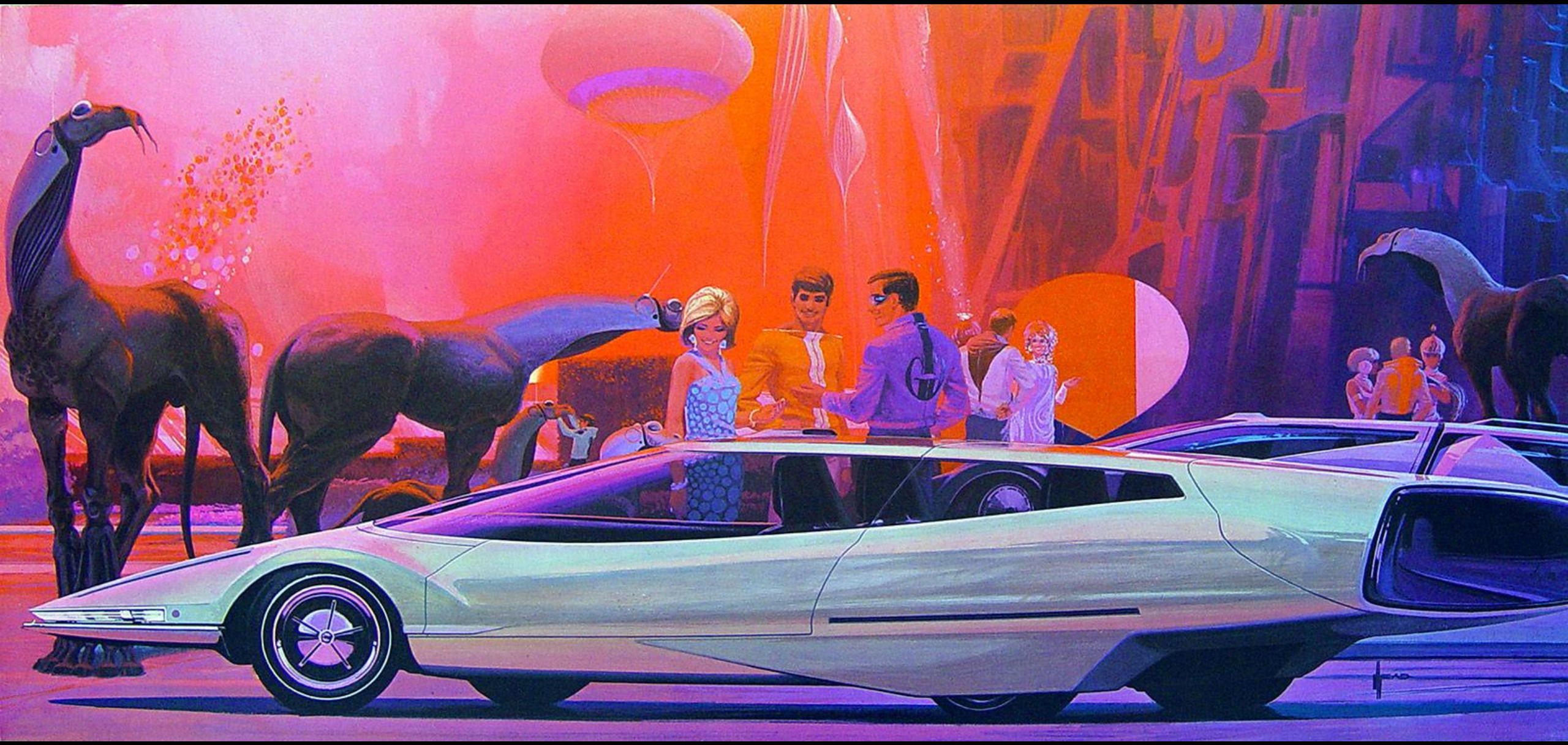
Generative Fill (Adobe Firefly)

Style Aligned Image Generation [Hertz et al. 2024]



Style Aligned Image Generation [Hertz et al. 2024]



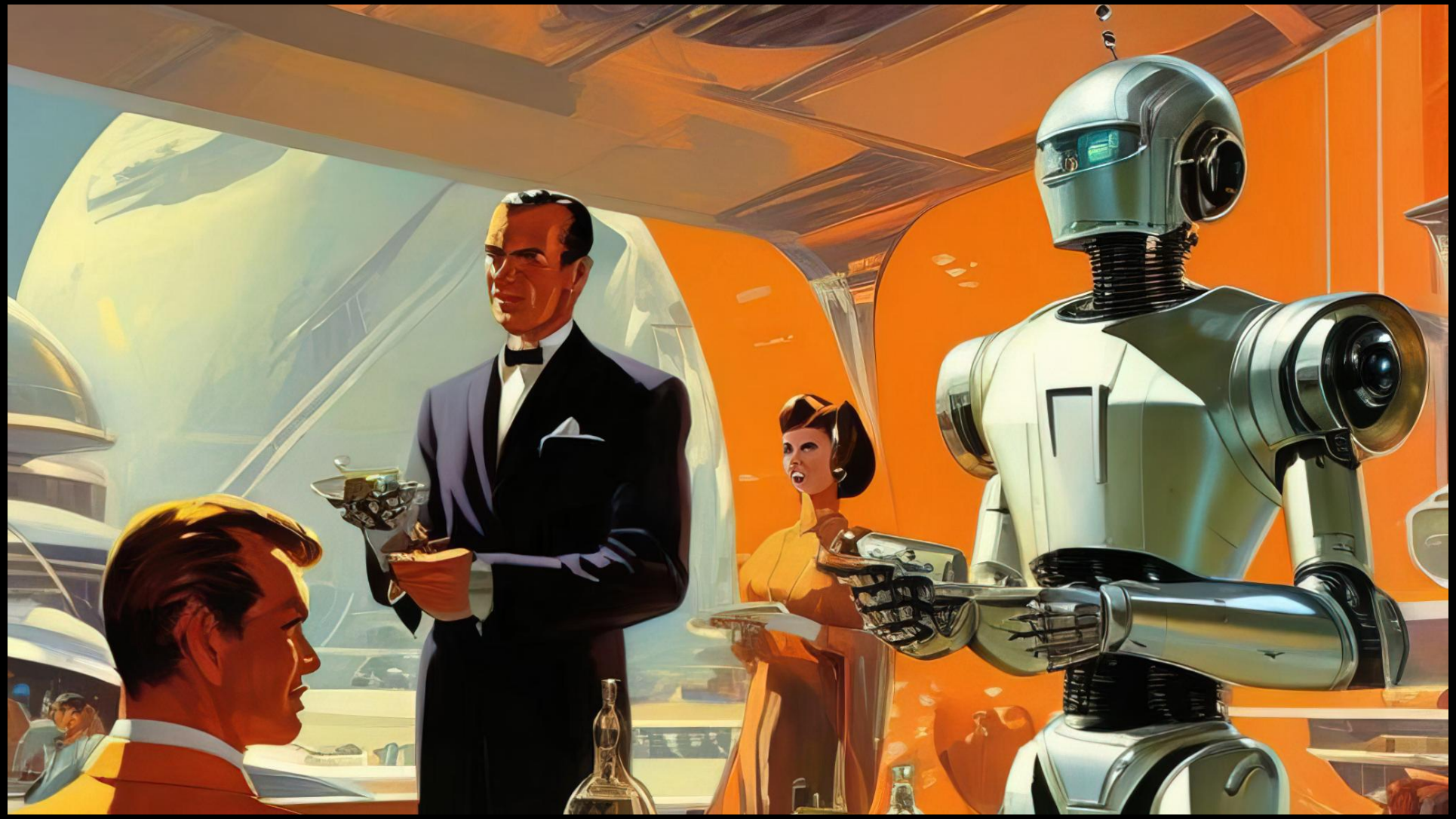


Syd Mead

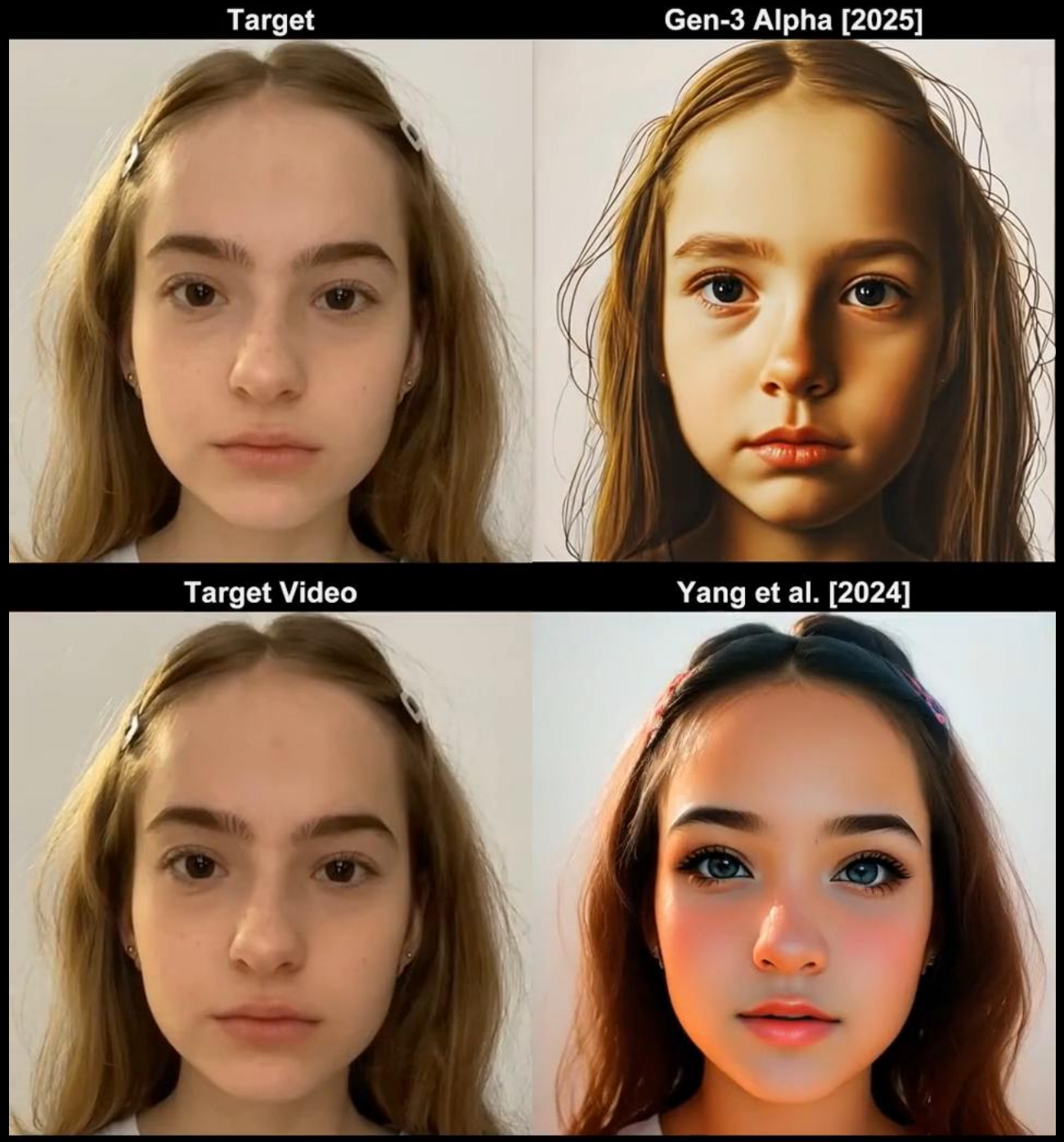
Syd Mead

“Syd Mead” (in reality)

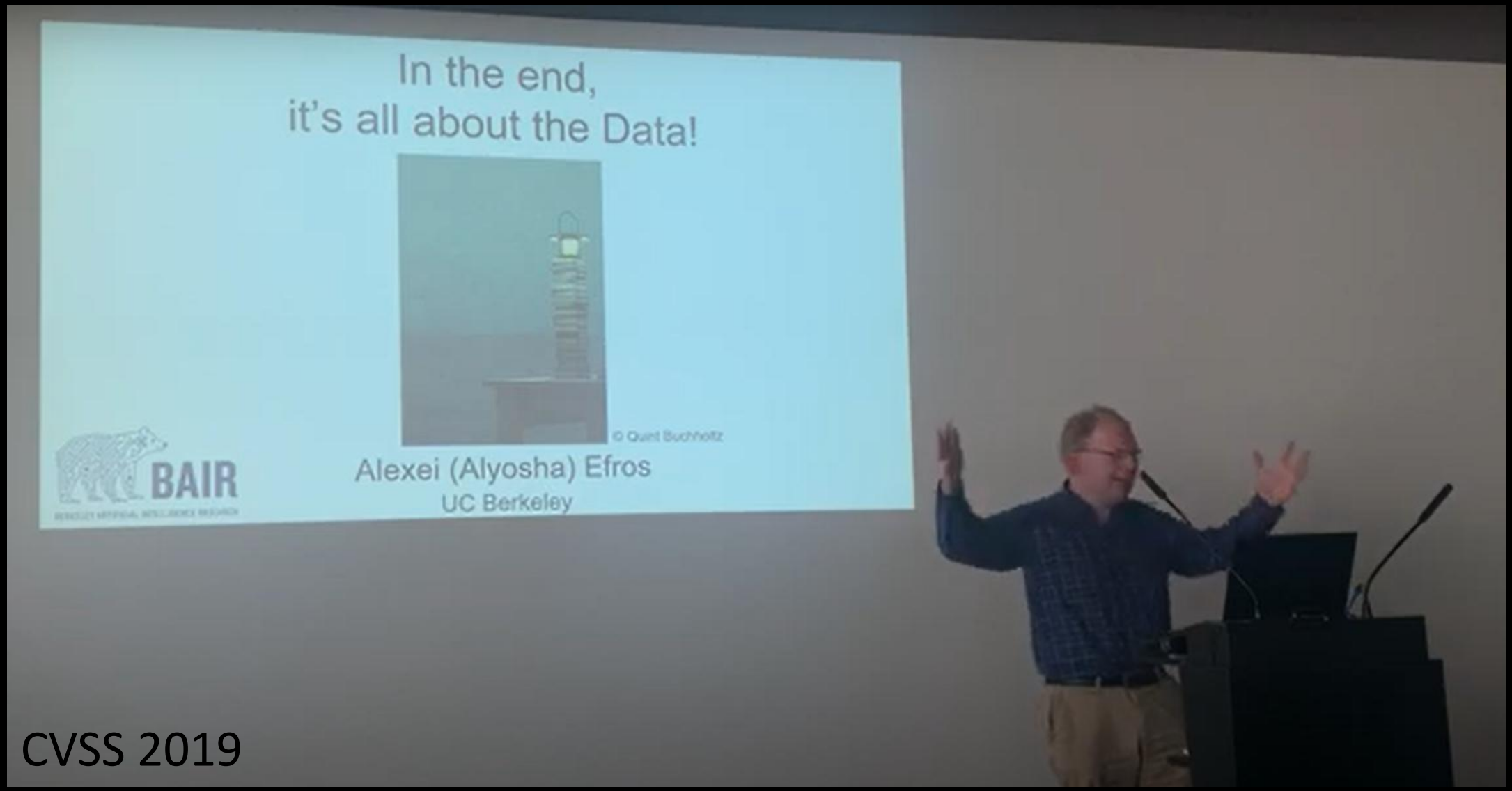
“Syd Mead” (synthetic)



“Syd Mead” (synthetic)

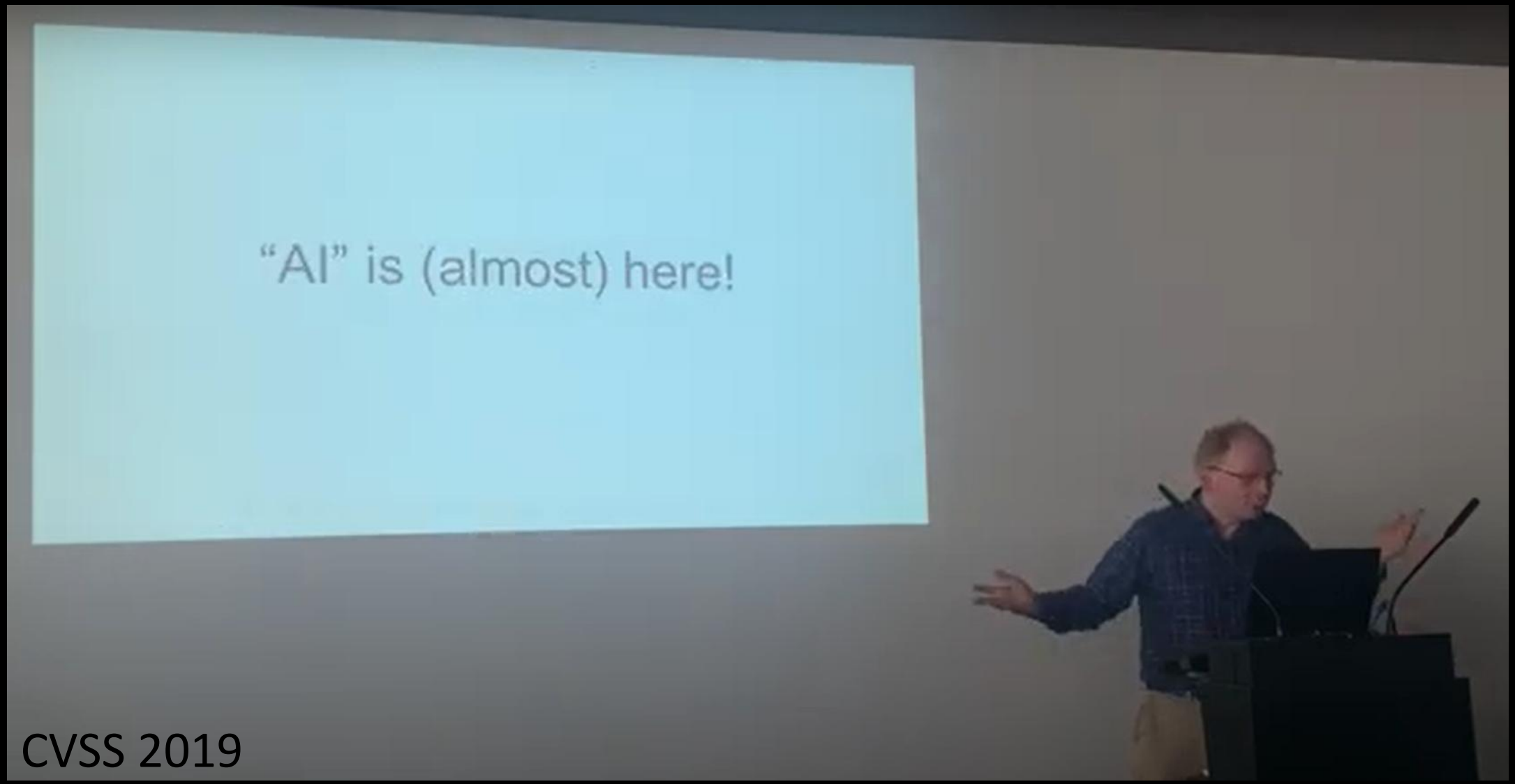

Target
Gen-3 Alpha [2025]
Target Video
Yang et al. [2024]

Artificial Intelligence



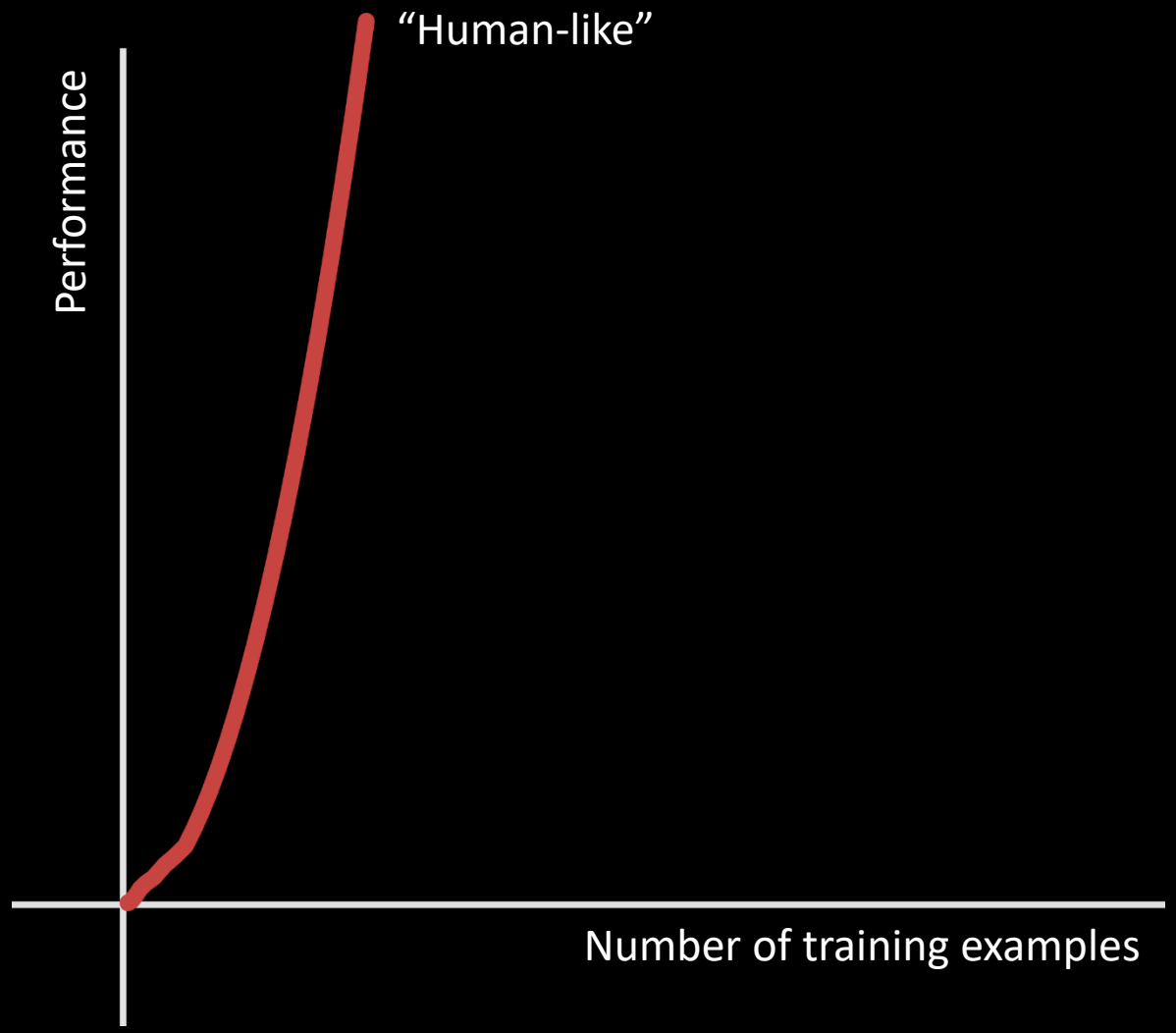
CVSS 2019

Artificial Intelligence

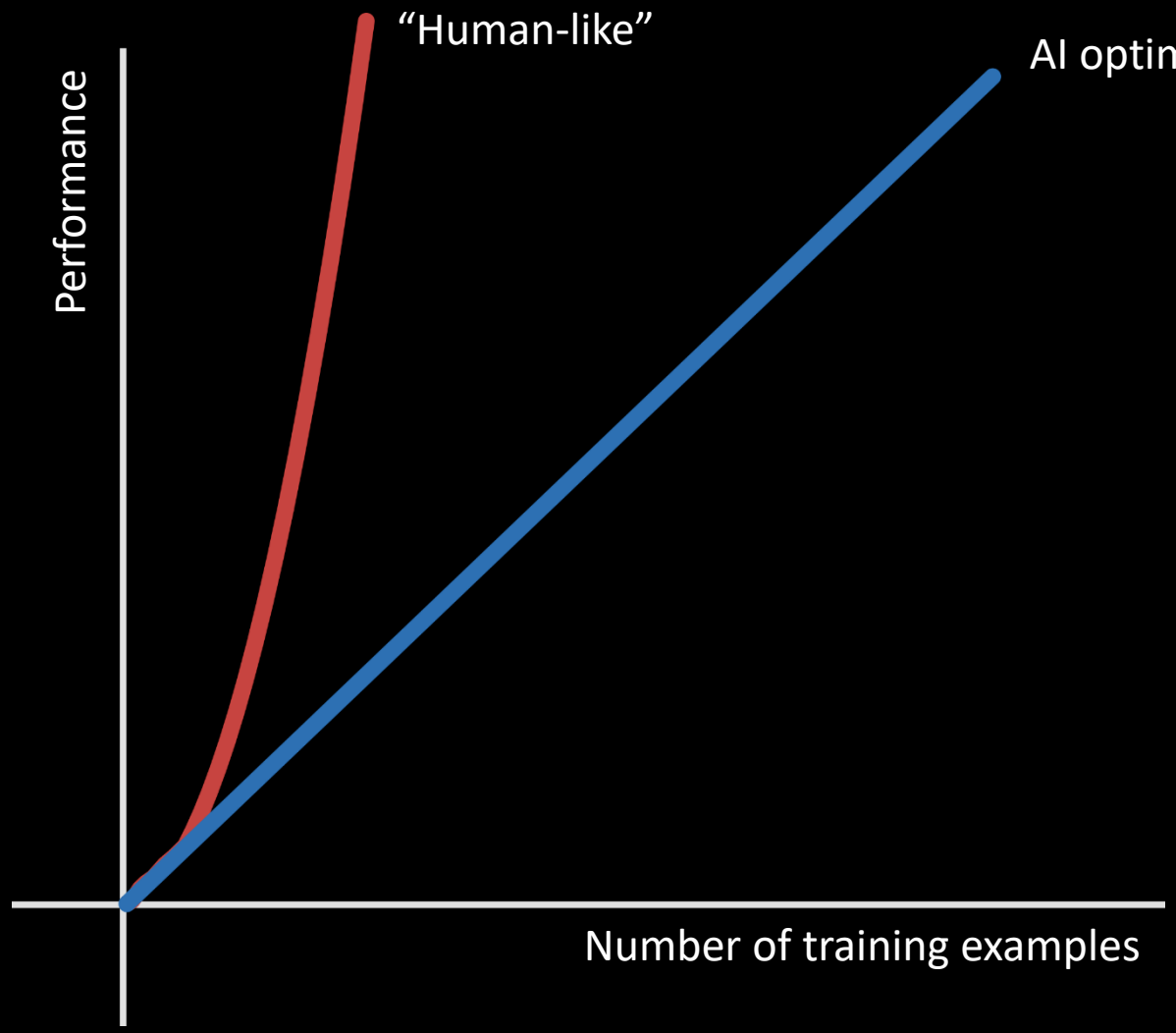


Artificial Intelligence

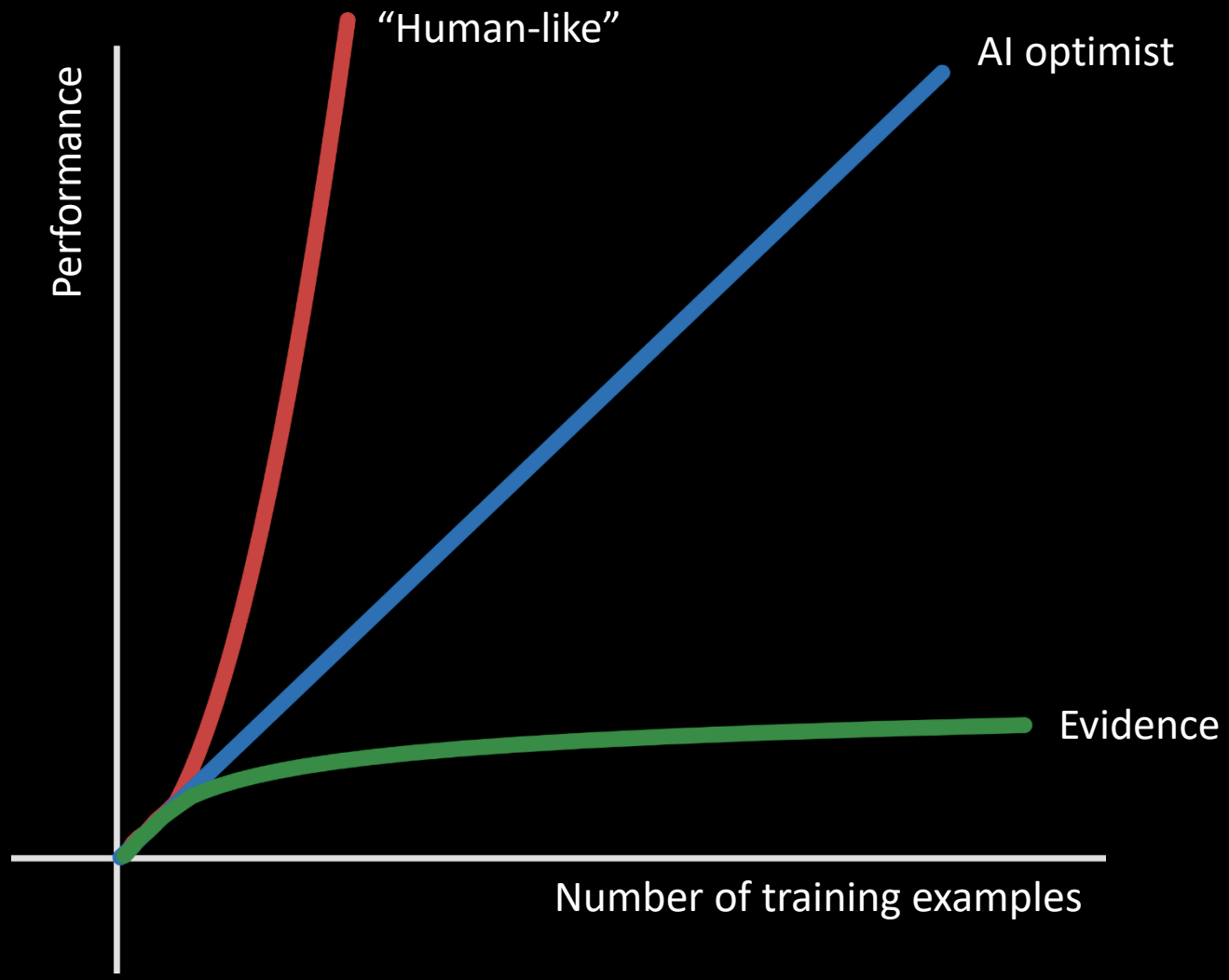
Udandarao et al.: No "Zero-Shot" Without Exponential Data: Pretraining Concept Frequency Determines Multimodal Model Performance, ICLR 2024.



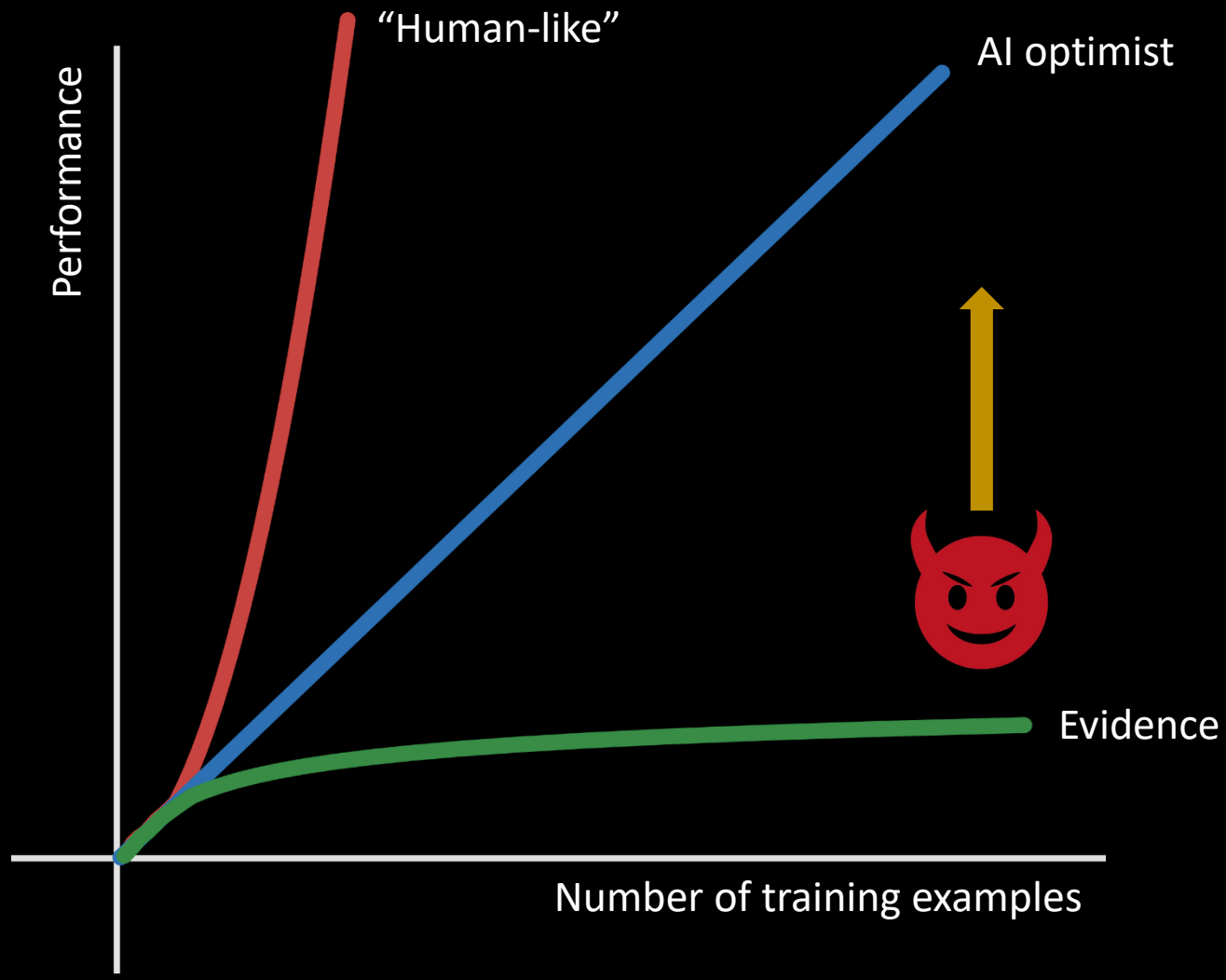
Udandarao et al.: No "Zero-Shot" Without Exponential Data: Pretraining Concept Frequency Determines Multimodal Model Performance, ICLR 2024.



Udandarao et al.: No "Zero-Shot" Without Exponential Data: Pretraining Concept Frequency Determines Multimodal Model Performance, ICLR 2024.



Udandarao et al.: No "Zero-Shot" Without Exponential Data: Pretraining Concept Frequency Determines Multimodal Model Performance, ICLR 2024.



Breaking Through GenAI's Quality Ceiling

Low-level textural details:

Source

Breaking Through GenAI's Quality Ceiling

Low-level textural details:

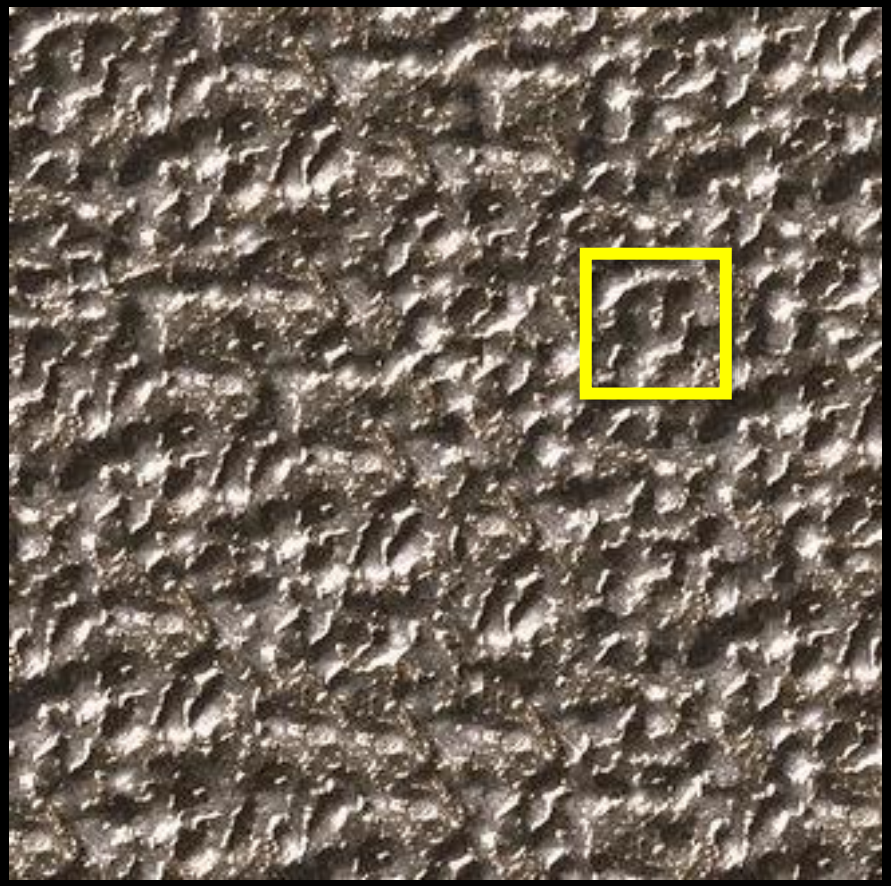
Source

Target

Breaking Through GenAI's Quality Ceiling

Low-level textural details:

Source

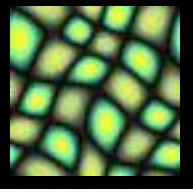


Target

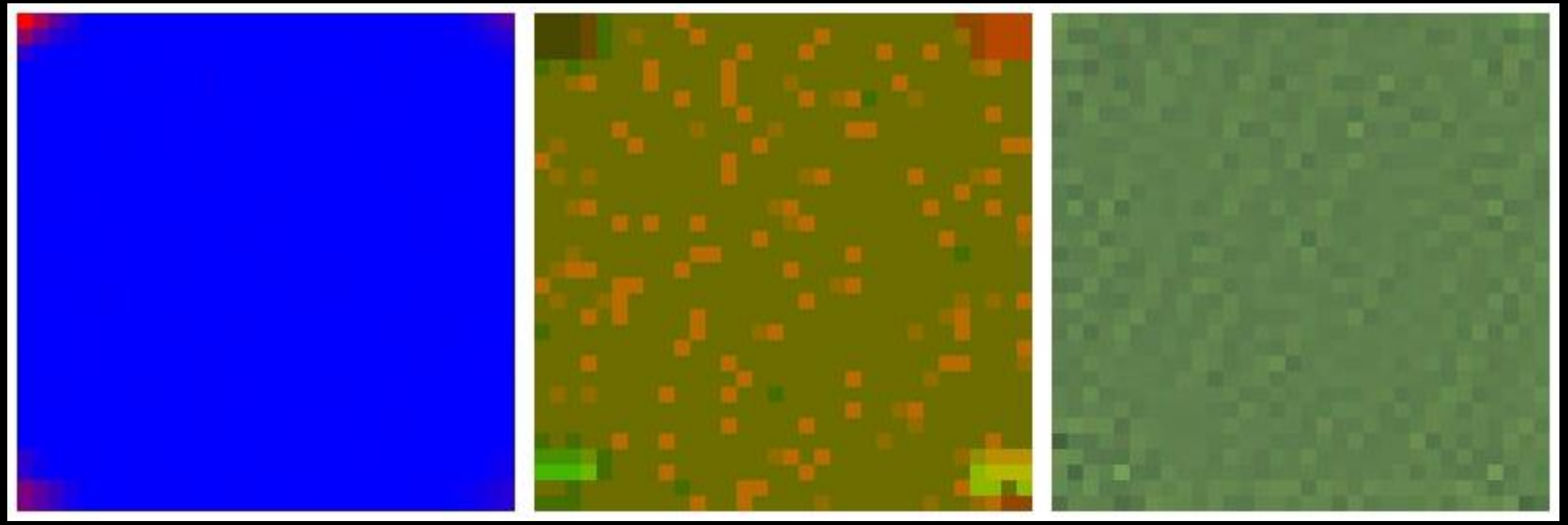
Algorithmic Solution: Patch-based Synthesis

Error

NNF



Output



Wexler et al. [2004], Kwatra et al. [2005], Kaspar et al. [2015], Fišer et al. [2016]

Breaking Through GenAI's Quality Ceiling

Seamless mosaic made of larger patches from the source:

Source

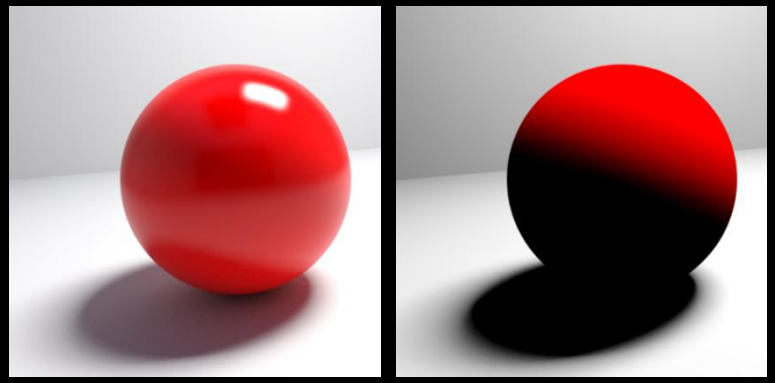
Target

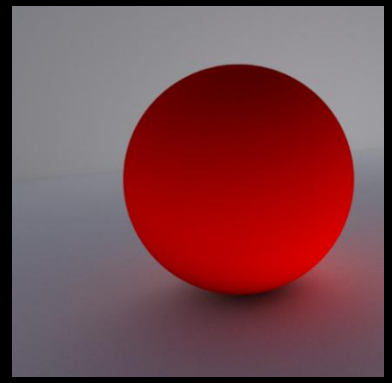
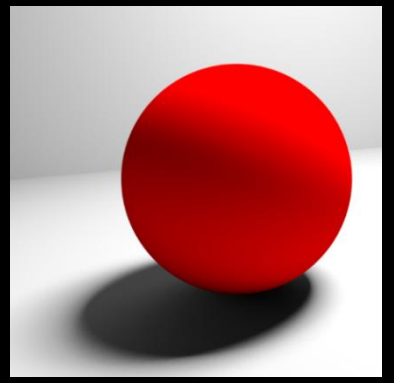
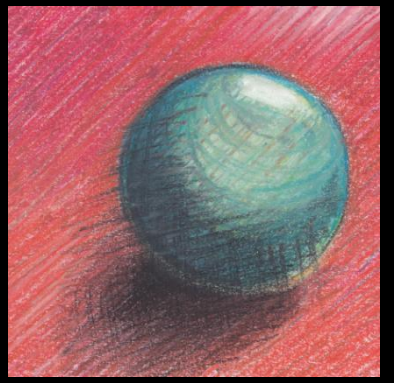
Generative Fill (Adobe Firefly)

Content-aware Fill (Adobe Photoshop)

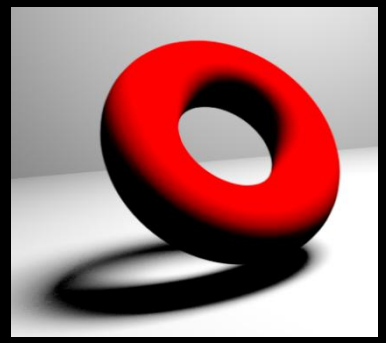
Breaking Through GenAI's Quality Ceiling

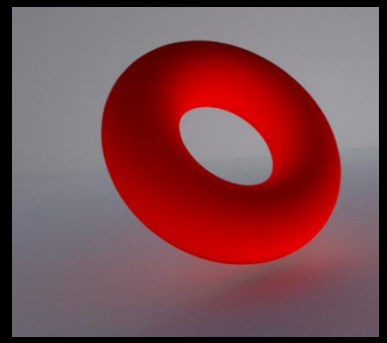
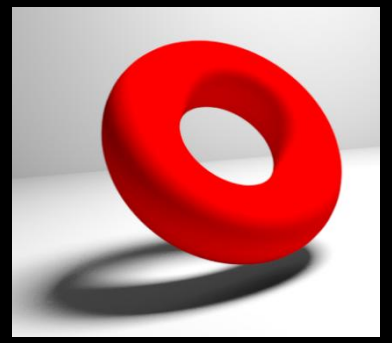
Source:



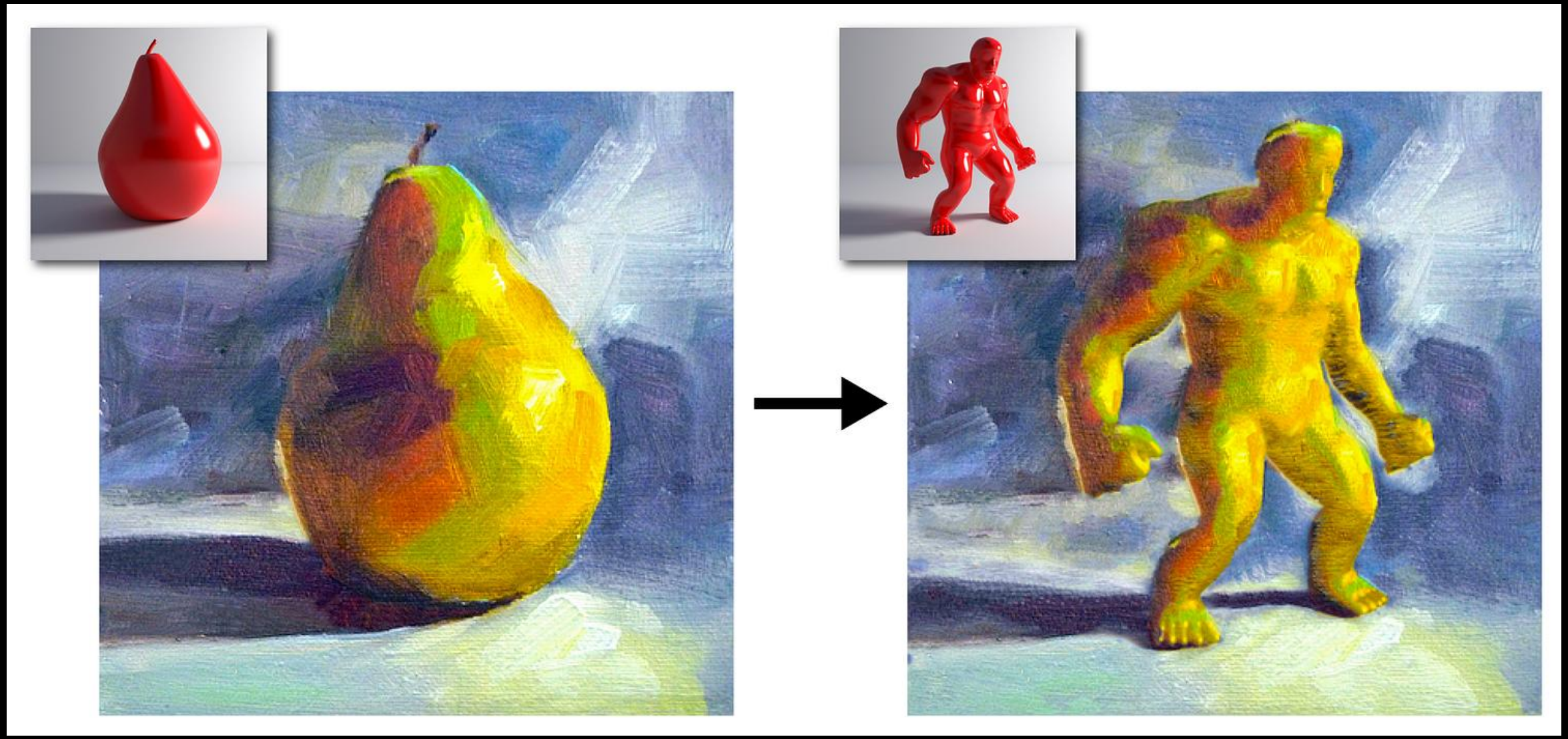


:


Target:

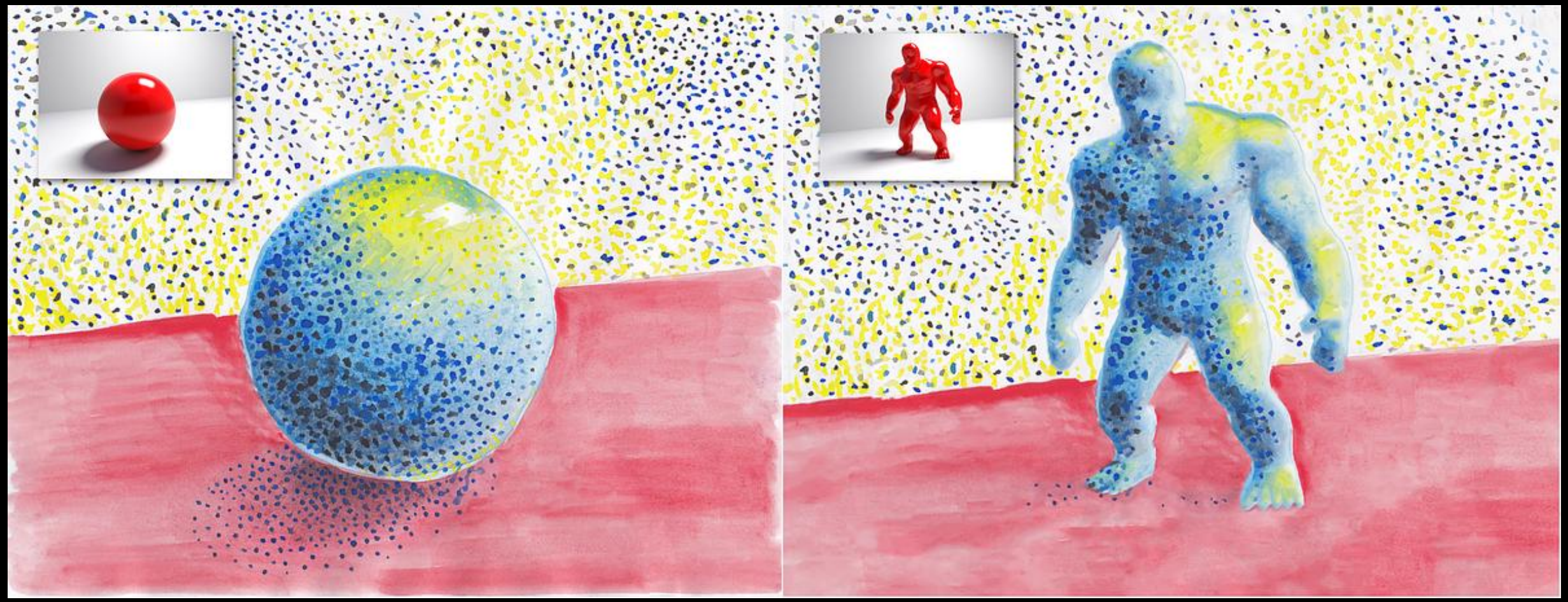





:
?

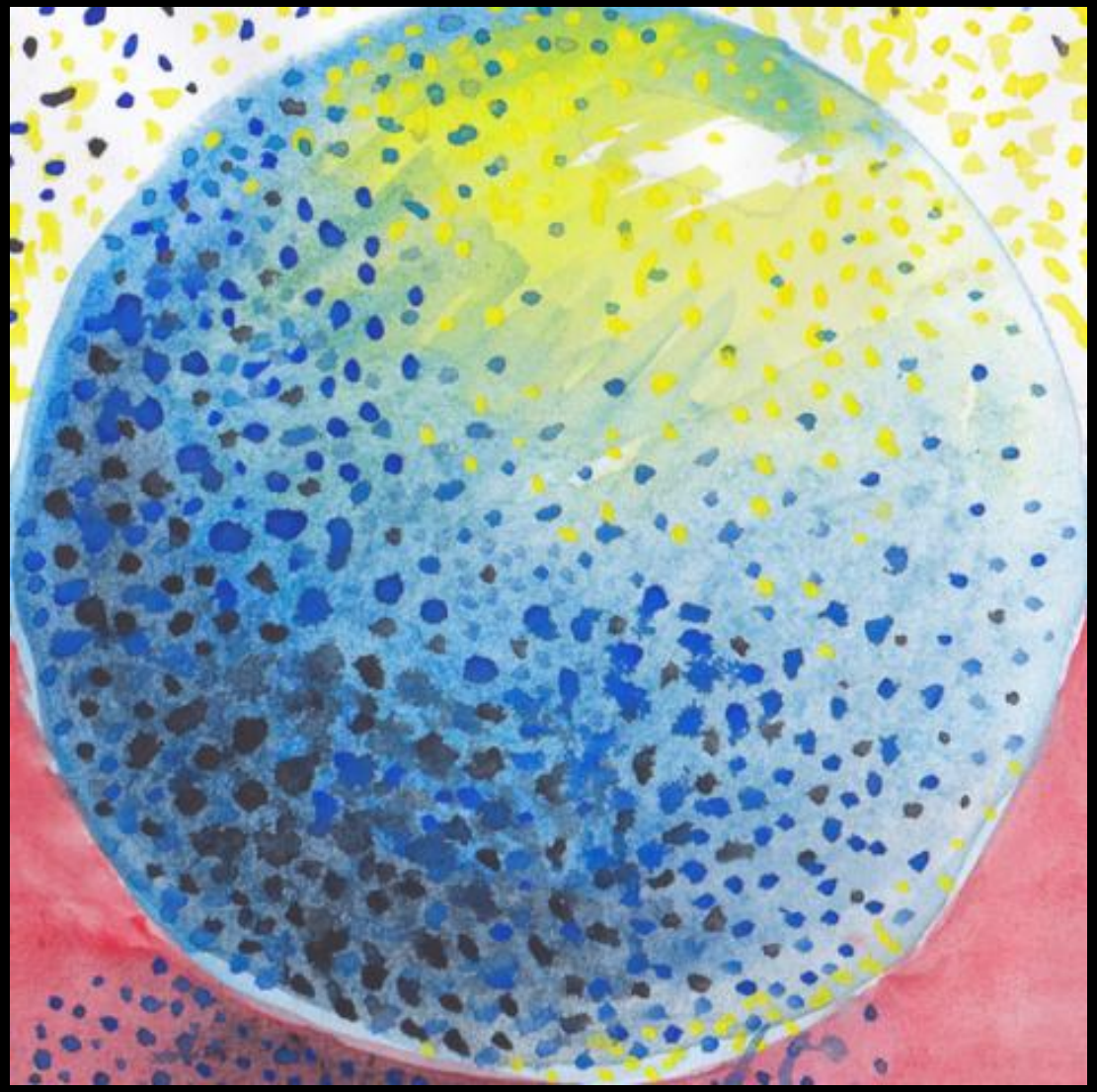
StyLit [Fišer et al. 2016]



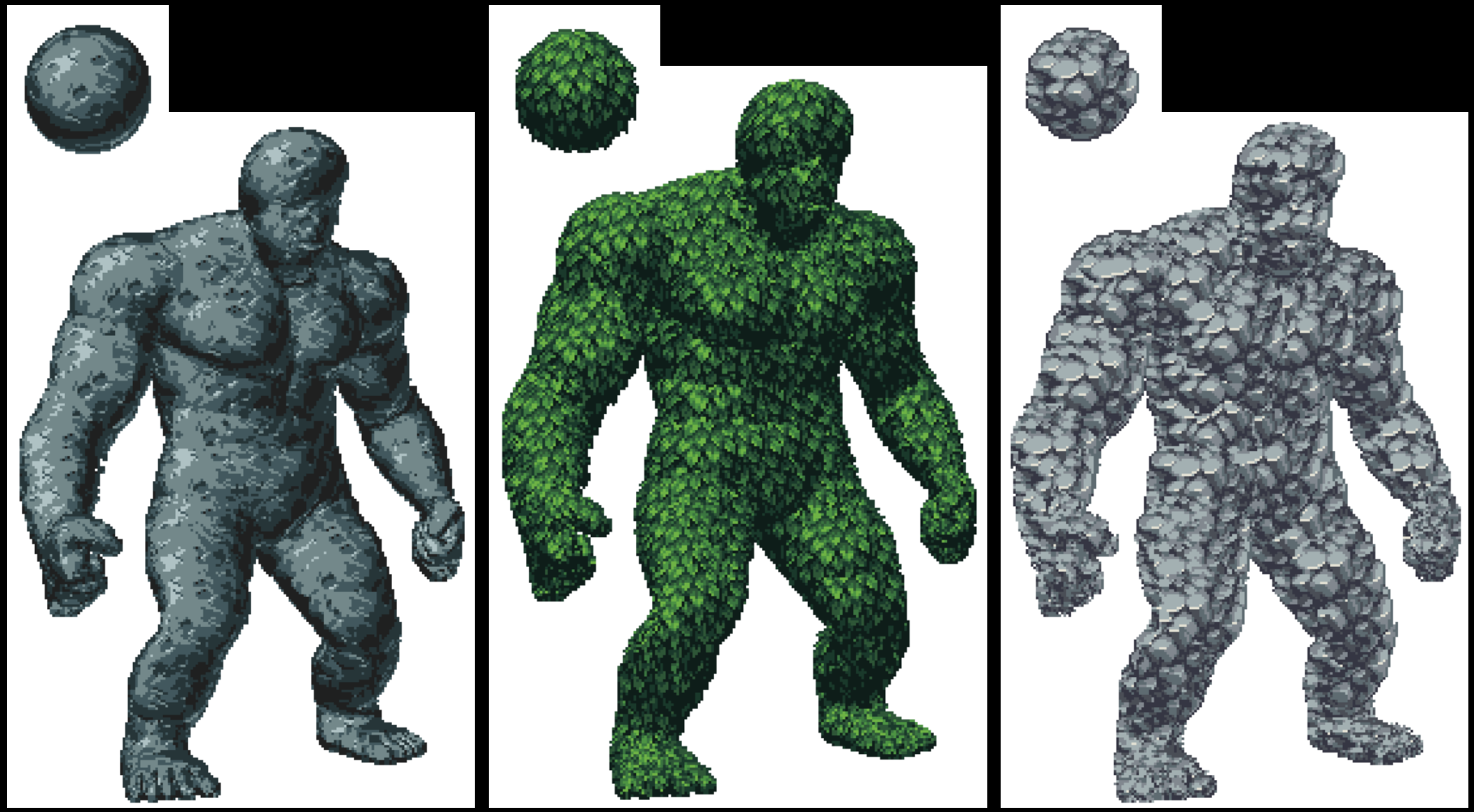
StyLit



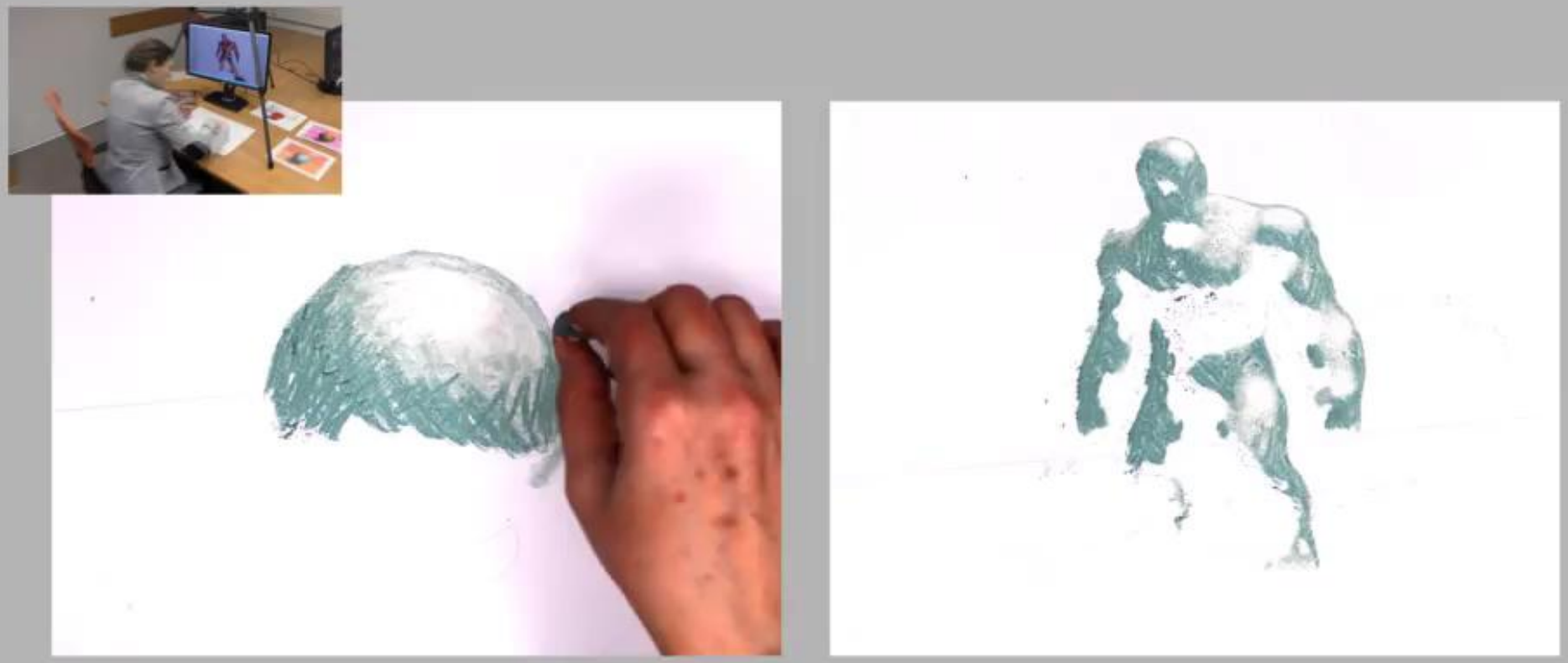
StyLit



StyLit

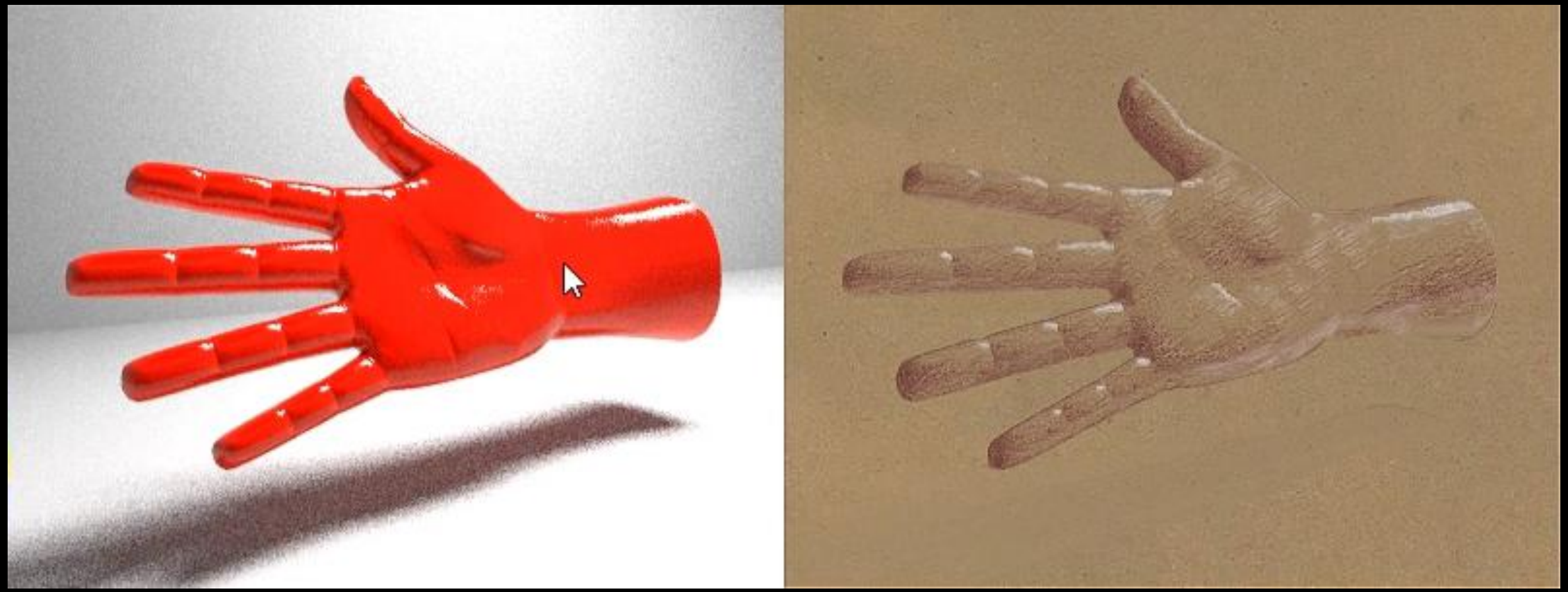


StyLit



Interactive Sessions (GeForce GTX 970)

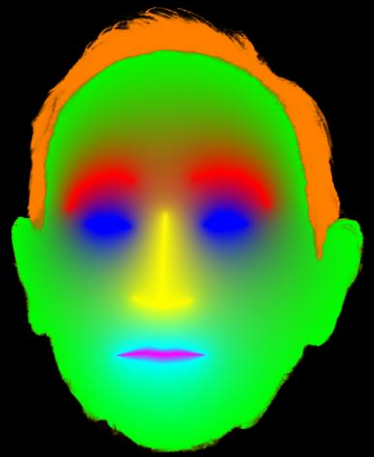
StyLit

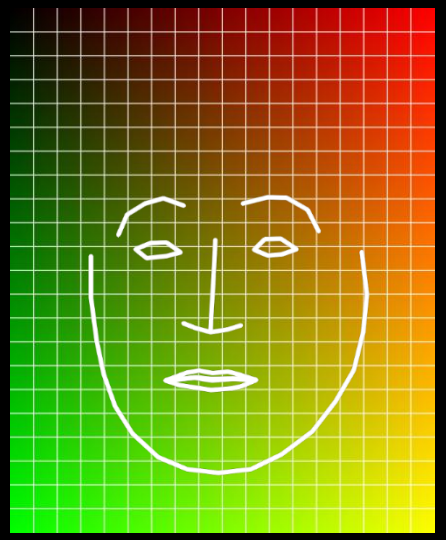


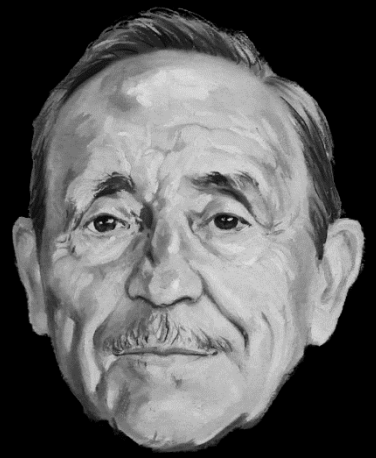
Interactive Sessions (GeForce GTX 970)

Breaking Through GenAI's Quality Ceiling

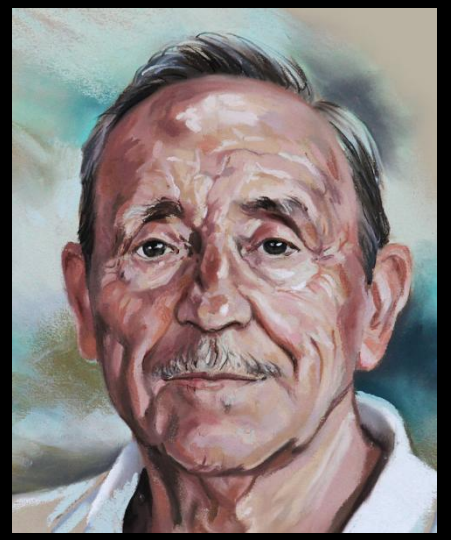
source



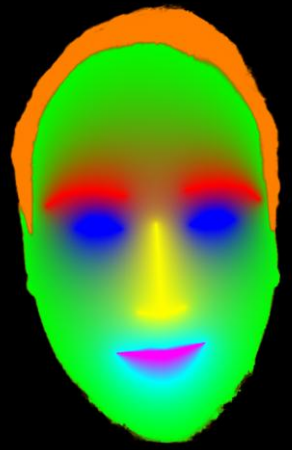


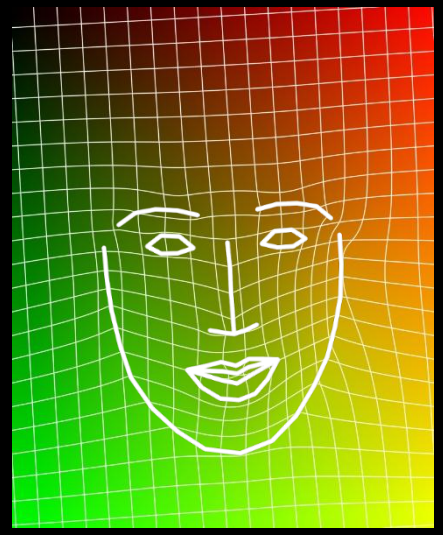


:



target





:

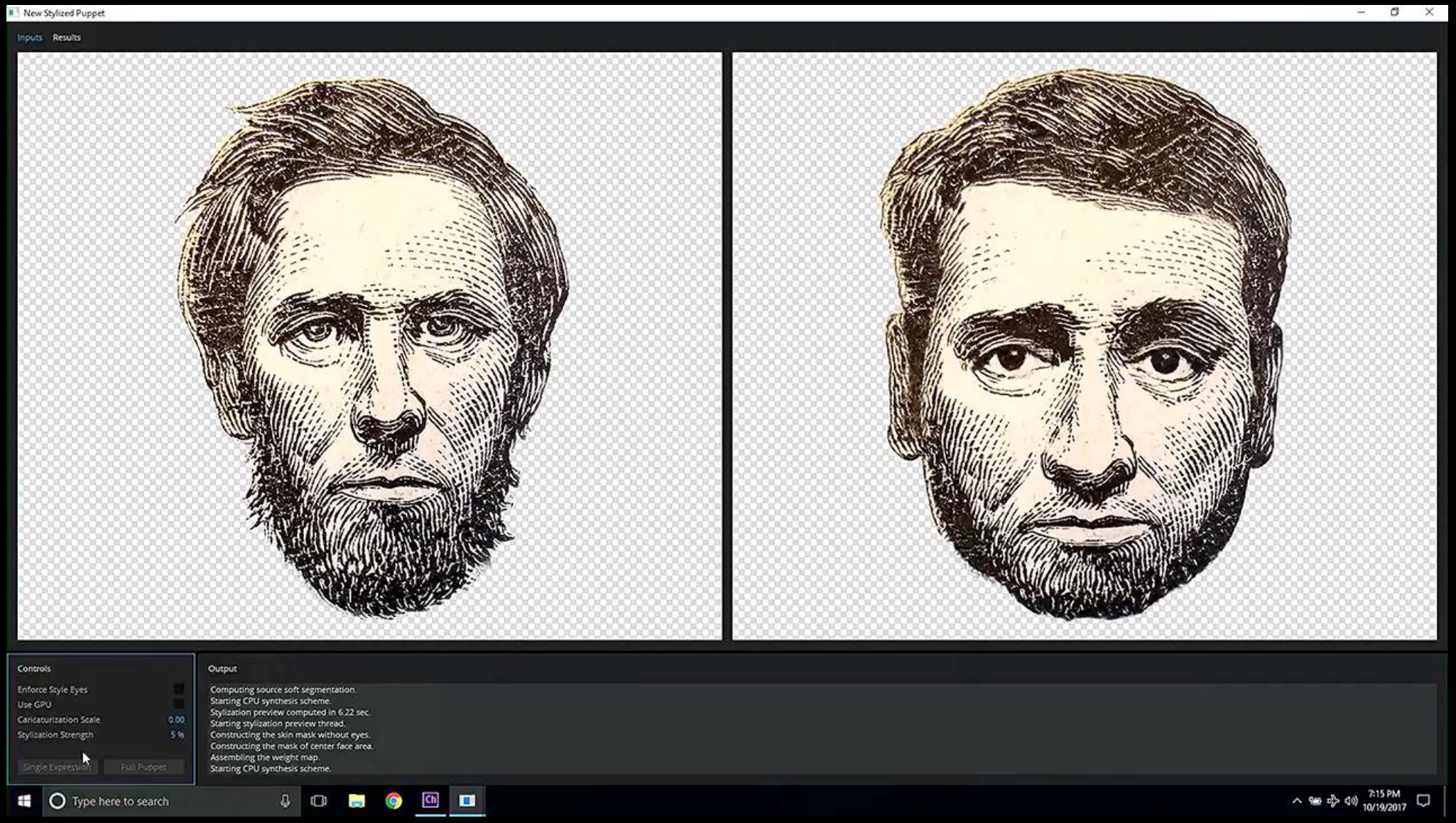
?

FaceStyle [Fišer et al. 2017]

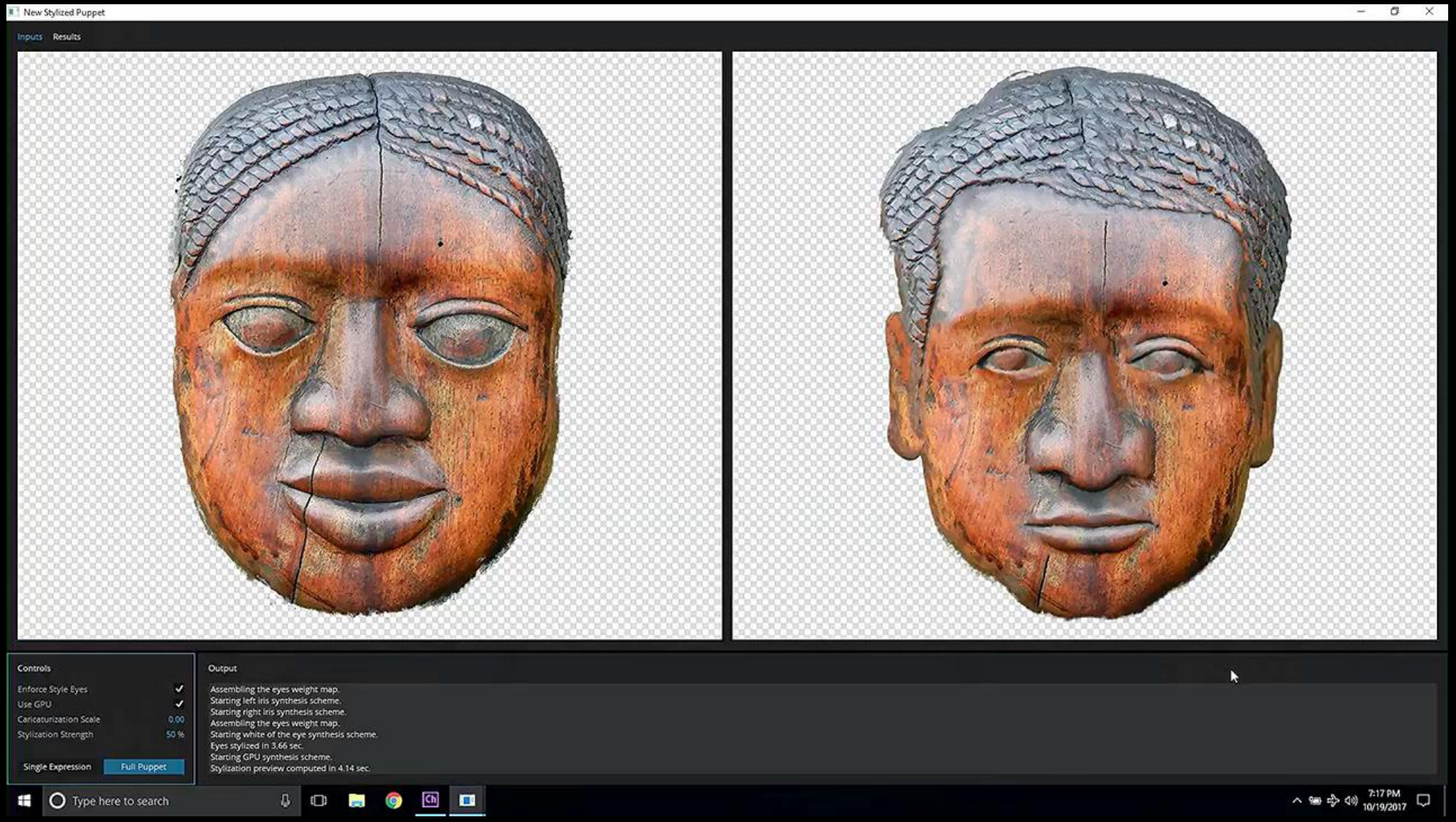
FaceStyle [Fišer et al. 2017]

Next Rembrandt with FaceStyle

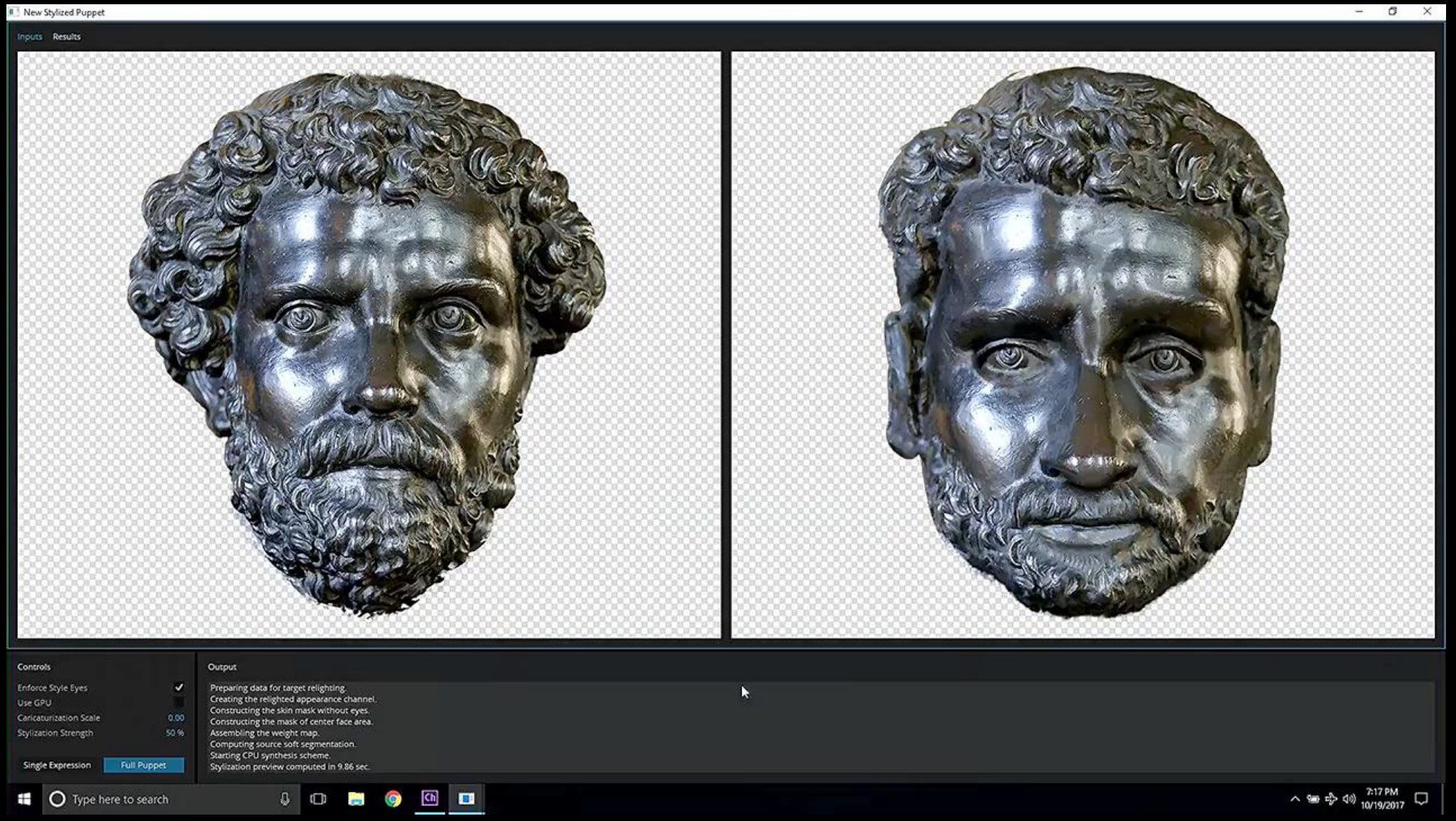
FaceStyle in Ch



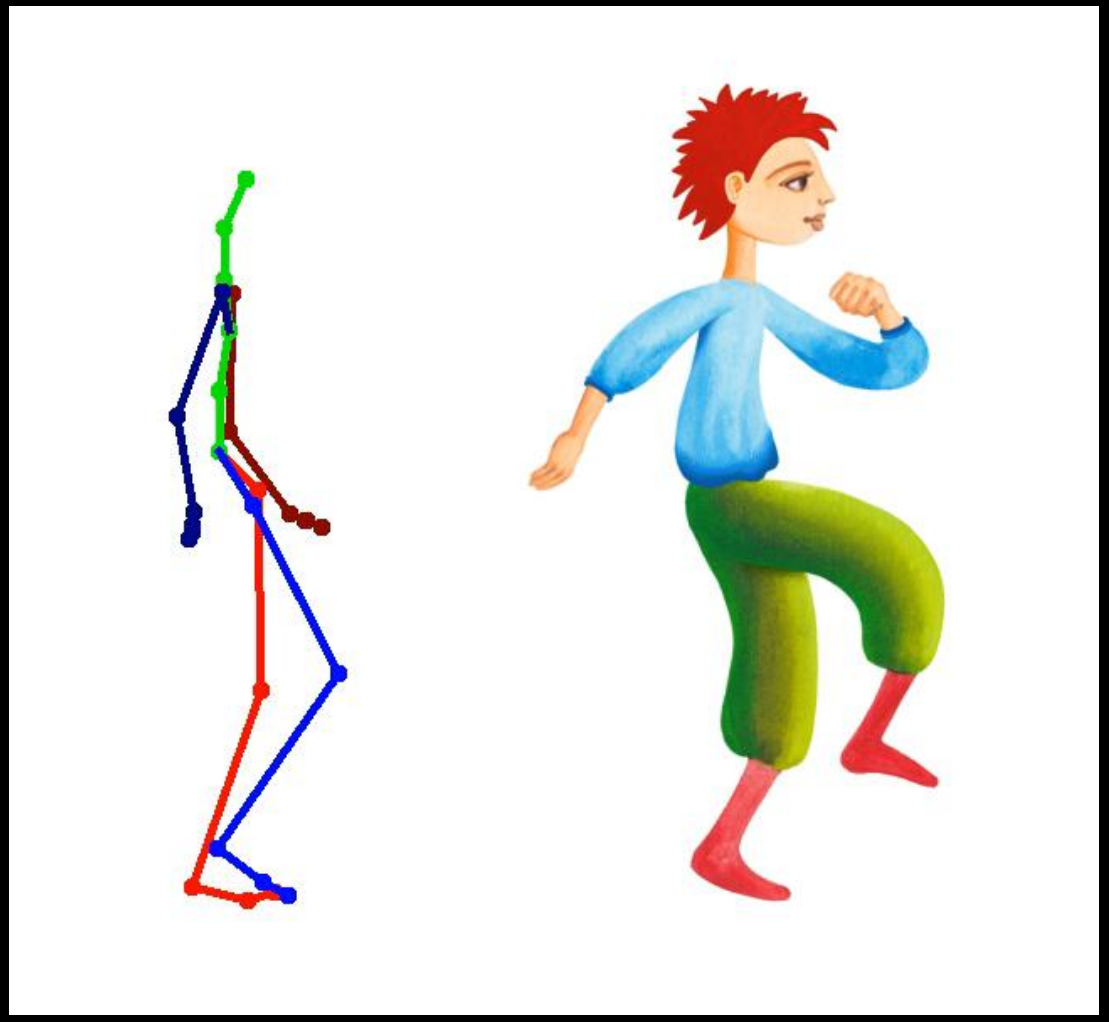
FaceStyle in Ch



FaceStyle in Ch

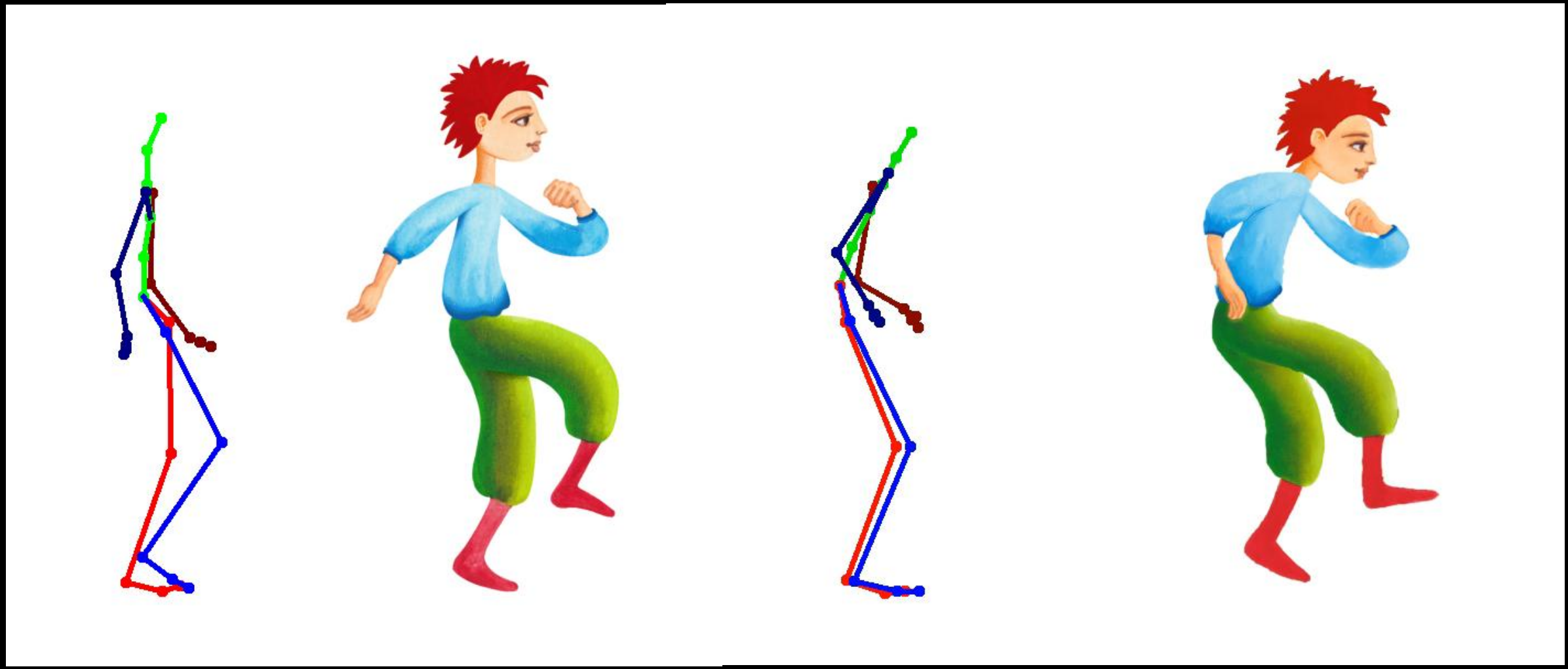


Style Transfer to Skeletal Animations



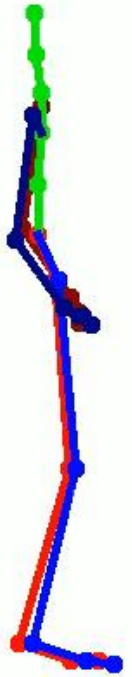
Source Style Exemplar

Style Transfer to Skeletal Animations

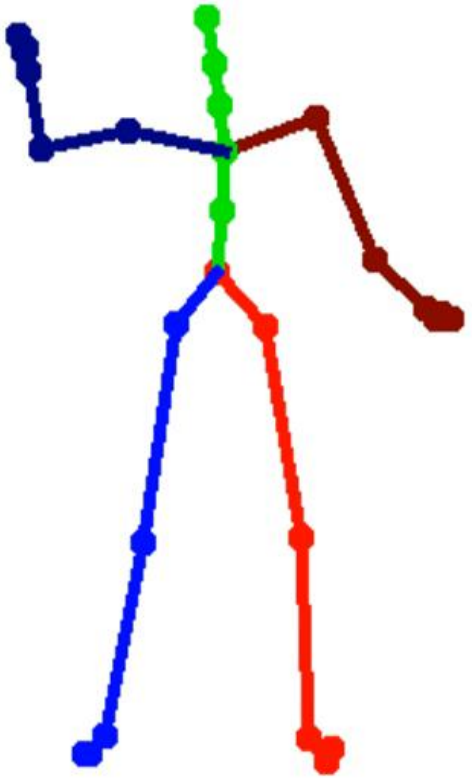


Source | Style Exemplar | Target | Output

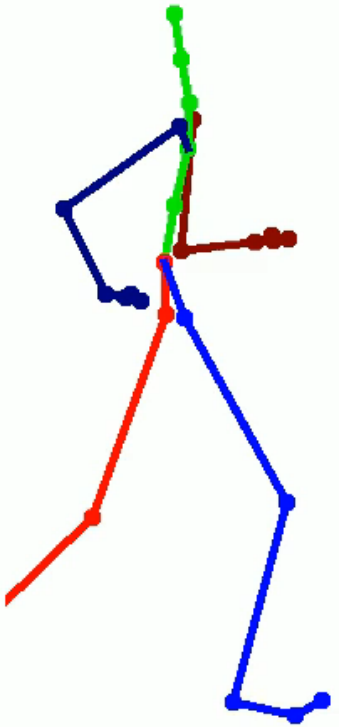
ToonSynth [Dvorožňák et al. 2018]



ToonSynth



ToonSynth



Stylizing Video by Example [Jamriška et al. 2019]

Keyframe

Stylized Keyframe

Remaining Frames to Stylize

MAUR film, Artwork: © Pavla Sýkorová

MAUR film, Artwork: © Václav Švankmajer


Artwork © Jakub Javora
Input Video
Input Keyframe #2
Synthesized Result

EbSynth

Change your video by editing one frame

EbSynth is VFX software for creative video transformations, retouching, and rotoscopy.

Start Now >

Watch Tutorial

http://ebsynth.com


INPUT VIDEO
KEYFRAME
RESULT

Thank you!