

画像生成AIについて

VONS
pictures

デジタル
コンテンツ
研究所

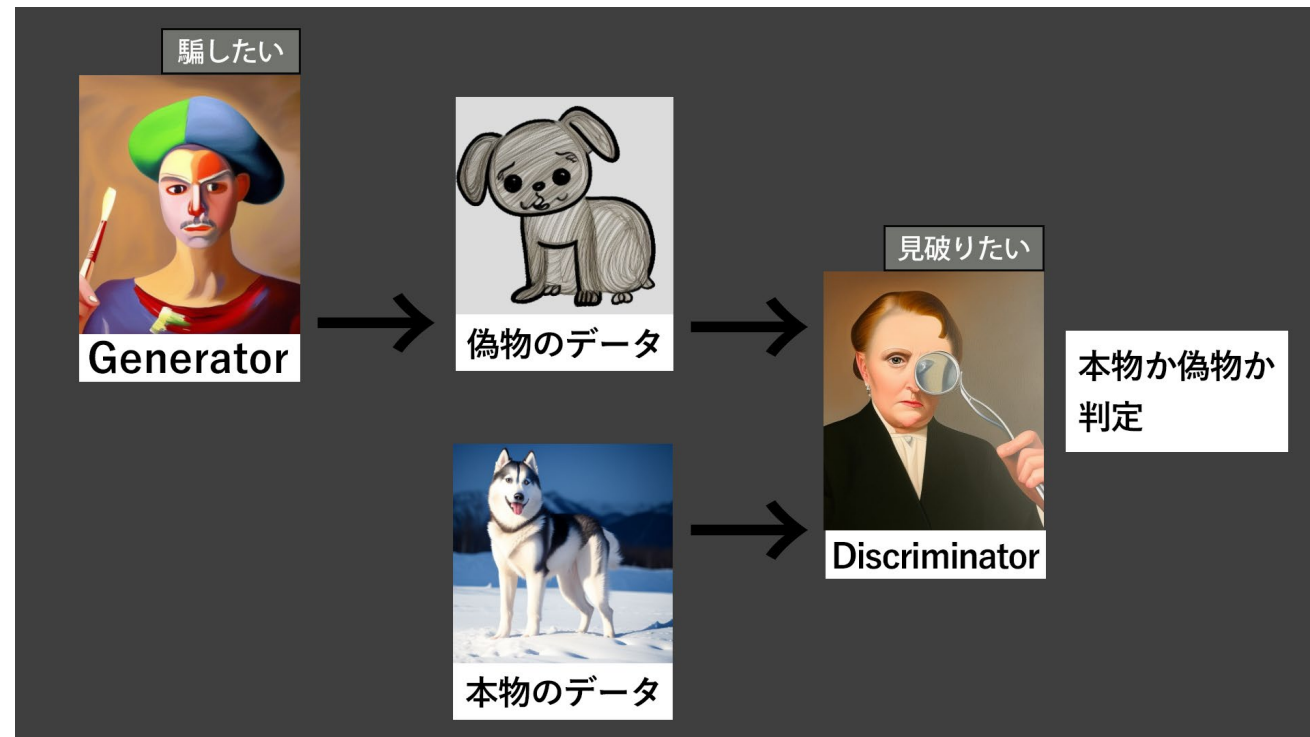
I.画像生成AI概論

従来の生成手法

2014年に発表された画像生成AIの一般的な手法であるGANは(Generative Adversarial Network)の略で**敵対的生成ネットワーク**と訳される。

GANは生成 (Generator) と識別 (Discriminator) の2つのネットワークで構成される。

GeneratorはDiscriminatorを騙すことができる本物そっくりの画像を生成するように学習し、Discriminatorはそうした見破るのが難しい偽物をより正確に見破れるように学習する。この2つが騙し合いを繰り返すことで画像生成の精度が上がる。



I.画像生成AI概論

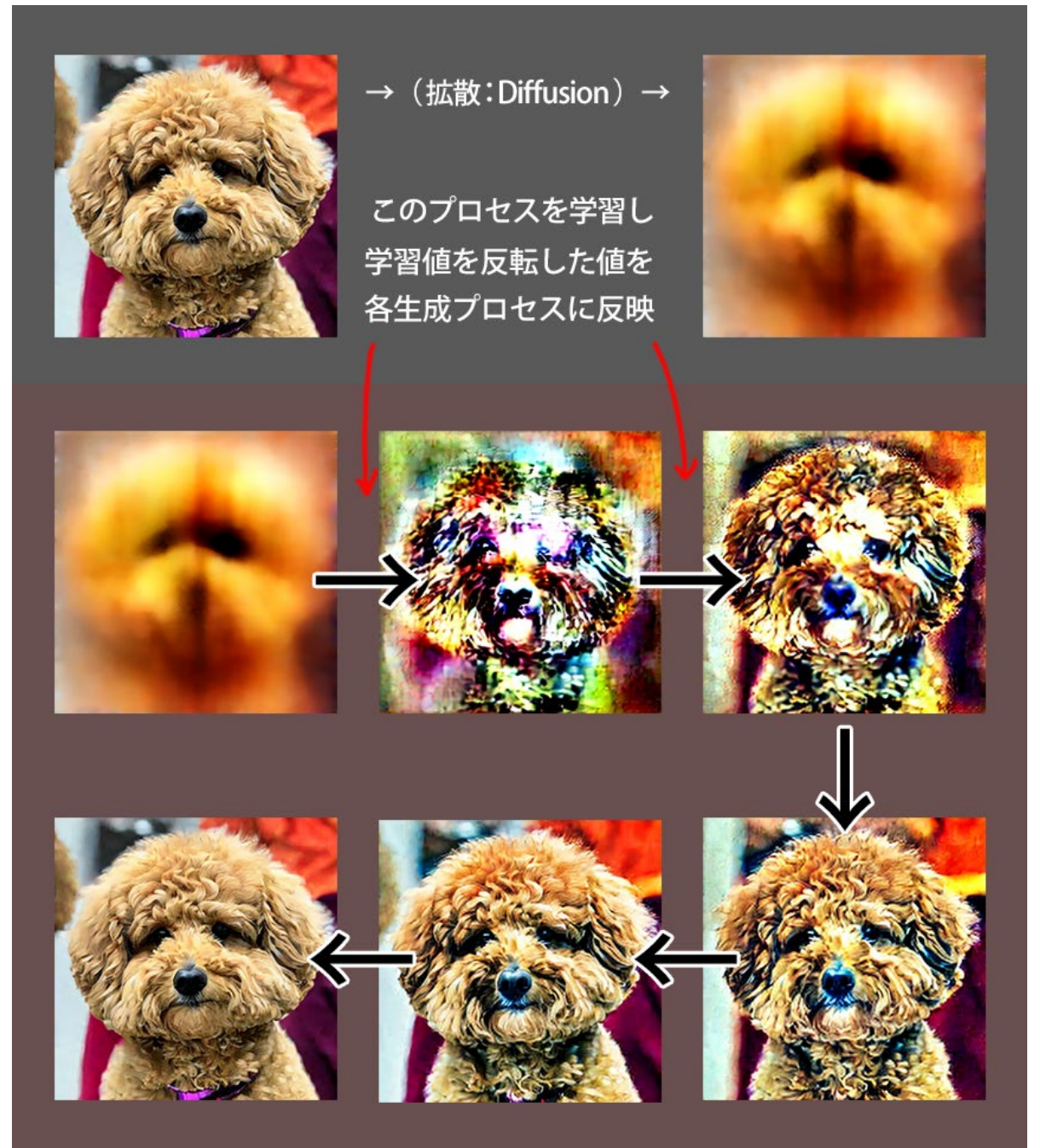
最先端の生成手法

一方ディフュージョンモデルは、ノイズを加えた画像から元の画像を復元するコンセプトを持つ。

段階的にノイズを加えていき、その各ステップをリバースプロセスすることで、各ステップにおける「元データに戻す」ためのパラメータを学習する。

ランダムな初期画像に対し、上述のパラメータを用いてノイズを段階的に取り除くことで、テキスト指示に従った新しい画像を生成する。

教師データである画像そのものはAIモデルの中には記憶されていない。



II.画像生成AIの実行環境

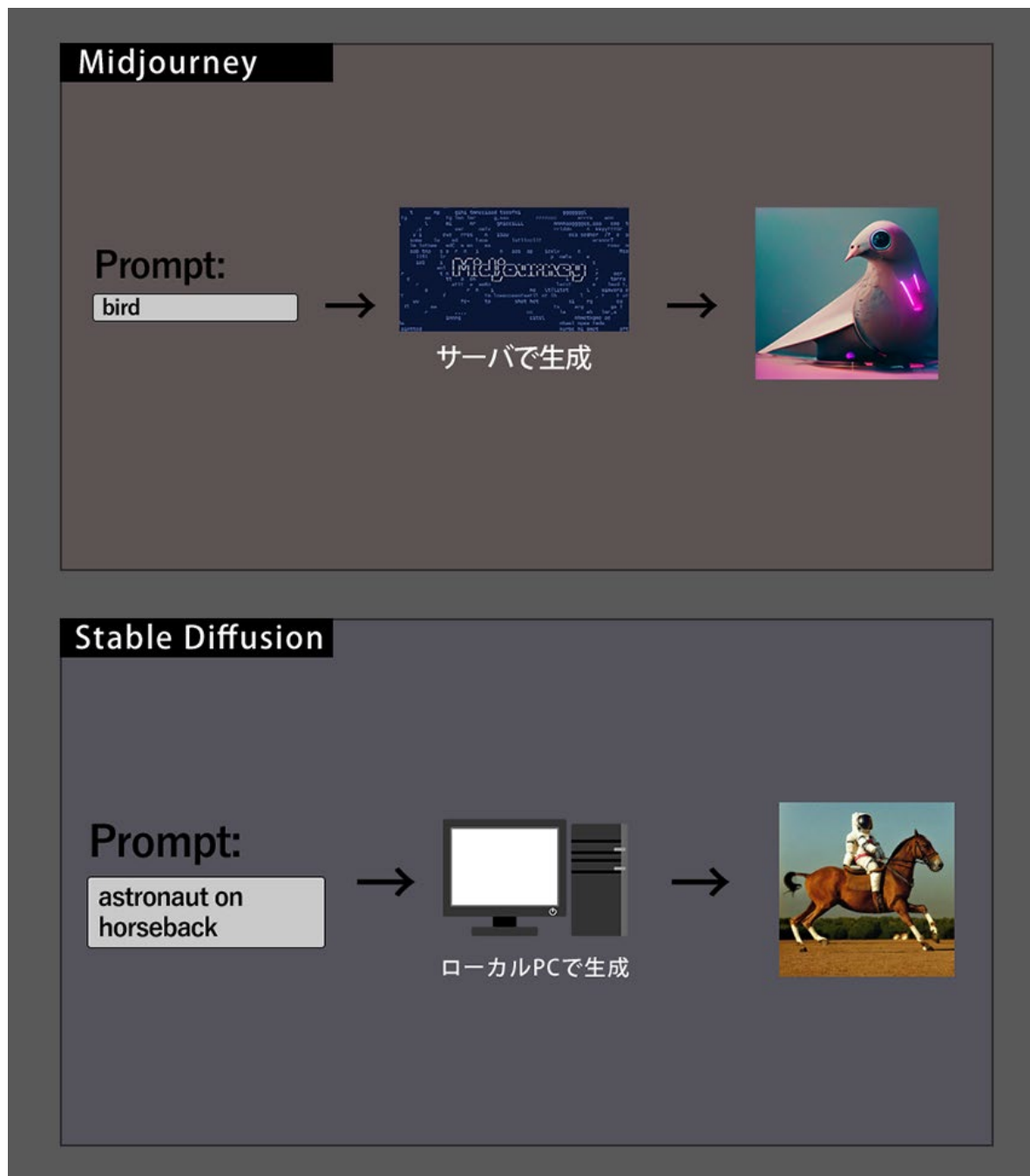
テキスト入力に基づくAI画像の実行環境は、インターネットサーバ経由で生成するか、ローカルPCで生成するかの2つを選択可能。

サーバ生成 Midjourney

Midjourneyは、テキストから画像を生成するAIを利用できるサービス。DiscordのBotを介してテキストを入力することで、サーバ側で画像が生成される。生成に用いられるAIモデルは非公開であり、その詳細は明らかにされていない。

PC生成 Stable Diffusion

StableDiffusionは、ディフュージョンモデルを用いた画像生成AI。公開されているモデルをダウンロードし、対応するプログラムを介して、ローカルPCのGPUを用いてテキストから画像を生成することが出来る。



III.Midjourneyと Stable Diffusionの画像比較

プロンプトの内容を同じにして生成される画像の
違いを比較してみる。

Midjourny Prompt:

App Icons, monochrome, no background, professional publicity photo of a clunky futuristic industrial vehicle with greebles, chipped paint weathered with rust and symmetrical tyres photorealistic portrait Lomography by Syd Mead hyper realistic octane render, 8k post processing is very detailed, moody lighting, Maya+V Ray +metal art+ extremely detailed, beautiful, unreal engine, ambient occlusion, subsurface scatter, metallic surfaces, radiosity, trending on Artstation, no icon, text, logo, ar 4:3

StableDiffusion Prompt:

App Icons, in the American southwest, professional publicity photo of a clunky futuristic industrial vehicle with greebles, chipped paint weathered with rust and symmetrical tyres photorealistic portrait Lomography by Syd Mead hyper realistic octane render, 8k post processing is very detailed, moody lighting, Maya+V Ray +metal art+ extremely detailed, beautiful, unreal engine, ambient occlusion, subsurface scatter, metallic surfaces, radiosity, trending on Artstation.

Midjourney



Stable Diffusion



III.Midjourneyと Stable Diffusionの画像比較

プロンプトの内容を同じにして生成される画像の
違いを比較してみる。

Midjourney Prompt:

a beautiful intricate fine art, luxury futuristic eyewear, full view of eyewear, product shot, studio lighting centered, analog photo::2, detailed::1.5, ZEISS, studio quality, 8k, 4k, uhd, photorealistic::2.5, detailed::2, transfer, no close up, flame, human, painting, drawing, sketch, camera, rendering, cropped, watermark, logo, barcode, UI, signature, text, label, error, title, stickers, markings, speech bubbles, lines, cropped, lowres, low quality, artifacts, ar 6:9

StableDiffusion Prompt:

A photograph of redshift style, a beautiful intricate fine art, luxury futuristic, eyewear, full view of eyewear, product shot, product cut, studio lighting centered,, (((analog photo))), ((detailed)), ZEISS, studio quality, 8k, 4k, uhd, (((photorealistic))), (((detailed))), transfer.

Negative prompt: close up, flame, human,, (((painting))), (((drawing))), (((sketch))), (((camera))), rendering, (((cropped))), (((watermark))), (((logo))), (((barcode))), (((UI))), (((signature))), (((text))), (((label))), (((error))), (((title))), stickers, markings, speech bubbles, lines, cropped, lowres, low quality, artifacts.

Midjourney



Stable Diffusion



III.Midjourneyと Stable Diffusionの画像比較

プロンプトの内容を同じにして生成される
画像の違いを比較してみる。

Midjourney Prompt:

inside, indoor, a futuristic sparse bedroom with large curved ceiling high windows looking out to a far future cyberpunk cityscape, flying drones outside, night time, cyberpunk neon lights, raining, analog photo::2, detailed::1.5, studio quality, 8k, 4k, uhd, photorealistic::2.5, detailed::2, transfer, --no fog, fire, backlight, painting, drawing, sketch, camera, rendering, cropped, watermark, logo, barcode, UI, signature, text, label, error, title, stickers, markings, speech bubbles, lines, cropped, lowres, low quality, artifacts, --ar 16:9

StableDiffusion Prompt:

A photograph of redshift style, inside, indoor, a futuristic sparse bedroom with large curved ceiling high windows looking out to a far future cyberpunk cityscape, flying drones outside, night time, cyberpunk neon lights, raining, (((analog photo))), ((detailed)), ZEISS, studio quality, 8k, 4k, uhd, photorealistic,((((detailed))), transfer.

Negative

prompt:((fog)),((fire)),((backlight)),((((painting))),((((drawing))),((((sketch))),((((camera))),((((rendering cropped))),((((watermark))), (((logo))), (((barcode))), (((UI))), (((signature))),(((text))),(((label))), (((error))), (((title))),stickers, markings, speech bubbles, lines,cropped, lowres, low quality, artifacts.

Midjourney

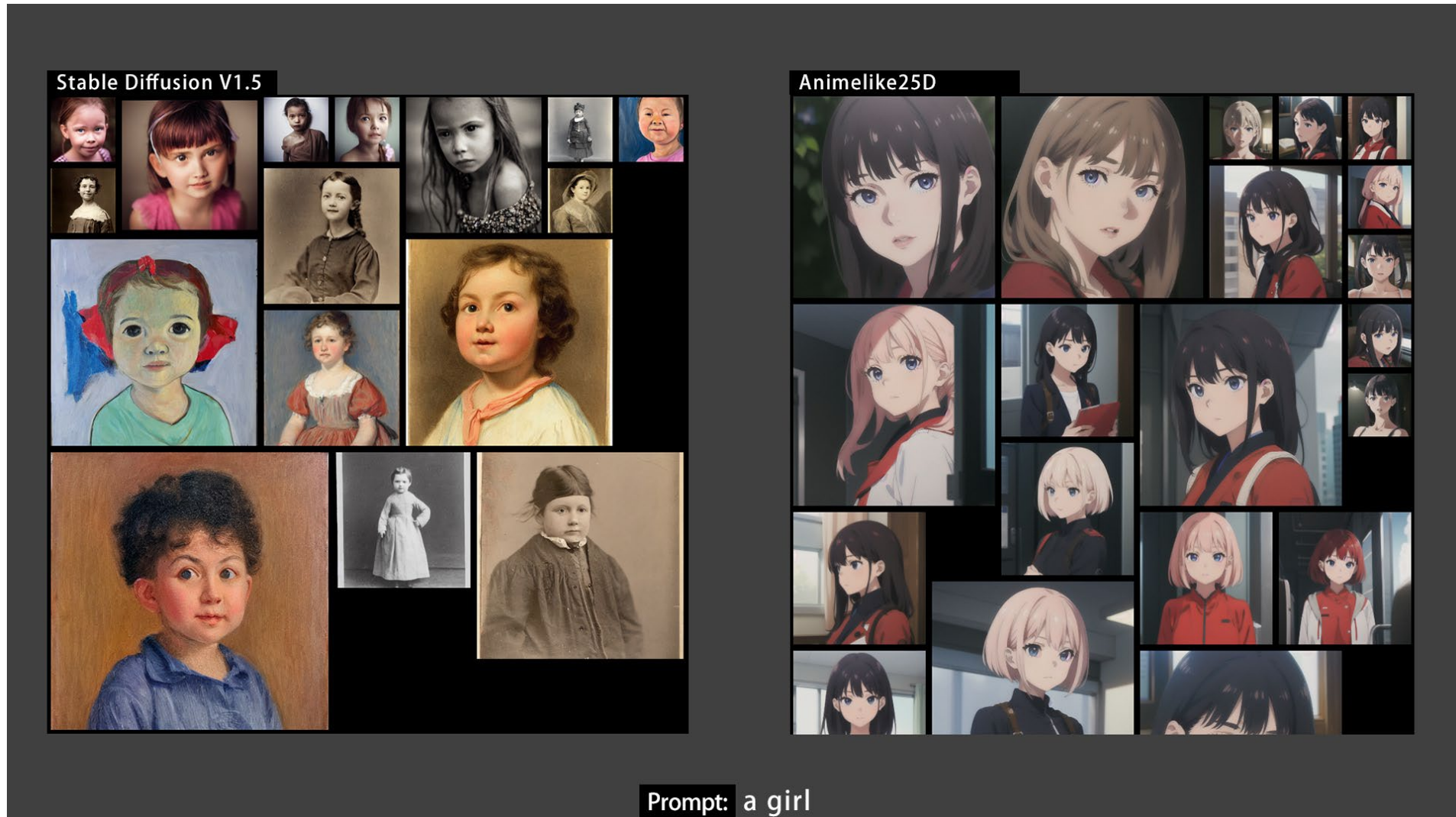


Stable Diffusion



IV. 画像生成の手法 モデル切り替え

前述のMidjourneyとStable Diffusionはモデルの違いにより出力される画像の傾向が異なる。下記右のモデルはアニメ画像に特化したモデルで、写真のような写実的画像はプロンプトで指定しても生成できない。



IV. 画像生成の手法 テキストtoイメージ(t2i)

テキストから画像への変換 (t2i) は、画像生成AIのタスクの1つであり、テキスト (プロンプト) によって目的の画像を指定する。例えば、"a girl"で人間の少女を認識させ、"a smiling girl"で笑顔に、"A photograph~"でスタイルを実写にするというように、AIが認識できるイメージの範囲を狭めていくことで目的の画像を得る。カンマやスペースで区切った単語の羅列でも文章でも良い。GANやディフュージョンモデルなど、さまざまなAIアルゴリズムがt2iタスクに使用されている。

Prompt: a girl,



Prompt: a smiling girl,



Prompt:

A photograph of a smiling girl,, ((analog photo)), (detailed), ZEISS, studio quality, 8k, 4k, uhd, (((photorealistic))), ((detailed)), transfer

Negative prompt: , ((painting)), ((drawing)), ((sketch)), ((camera)), ((rendering)),(((cropped))), (((watermark))), ((logo)), ((barcode)), ((UI)), ((signature)), ((text)), ((label)), ((error)), ((title)), stickers, markings, speech bubbles, lines, cropped, lowres, low quality, artifacts



Prompt:

A photograph of A photograph of best quality, ultra high res, (photorealistic:1.4), 1girl, ██████████, ((analog photo)), (detailed), ZEISS, studio quality, 8k, 4k, uhd, (((photorealistic))), ((detailed)), transfer, ((analog photo)), (detailed), ZEISS, studio quality, 8k, 4k, uhd, (((photorealistic))), ((detailed)), transfer

Negative prompt: ██████████ paintings, sketches, (worst quality:2), (low quality:2), (normal quality:2), lowres, normal quality, ((monochrome)), ((grayscale)), skin spots, acnes, skin blemishes, age spot, glans, ((painting)), ((drawing)), ((sketch)), ((camera)),((rendering)),(((cropped))), (((watermark))), ((logo)), ((barcode)), ((UI)), ((signature)), ((text)), ((label)), ((error)), ((title)), stickers, markings, speech bubbles, lines, cropped, lowres, low quality, artifacts, ((painting)), ((drawing)), ((sketch)), ((camera)), ((rendering)),(((cropped))), (((watermark))), ((logo)), ((barcode)), ((UI)), ((signature)), ((text)), ((label)), ((error)), ((title)), stickers, markings, speech bubbles, lines, cropped, lowres, low quality, artifacts



IV. 画像生成の手法 イメージtoイメージ(i2i)

テキストから画像を生成するt2iでは、画面レイアウトをプロンプトで細かくコントロールすることが困難な場合がある。そこで、既存の画像をベースに新たな画像を生成する手法として、イメージtoイメージ (i2i) がある。i2iは画像のスタイルや内容を変換することが可能で例えば、アートスタイルの変換や、画像内のオブジェクトや領域を識別して別々に扱うことが出来る。前述のモデル切り替えによるテキストの効き具合は、t2iよりも弱く、強くするとレイアウトが崩壊する。

original



Prompt:

A photograph of beautiful young boy, bustup angle, closeup,, ((analog photo)), (detailed), ZEISS, studio quality, 8k, 4k, uhd, (((photorealistic))),((detailed)), transfer, dribble, precisionism, associated press photo, award-winning

Negative prompt [redacted] ((painting)), ((drawing)), ((sketch)), ((camera)), ((rendering)), mutilated, disfigured, unnatural, bad anatomy, morbid, too many fingers, deformed palm, deformed arm, deformed limbs, extra limbs, missing limbs, unnatural pose, (((cropped))), (((watermark))), ((logo)), ((barcode)), ((UI)), ((signature)), ((text)), ((label)), ((error)), ((title)), stickers, markings, speech bubbles, lines, cropped, lowres, low quality, artifacts



Prompt:

A photograph of beautiful 25 years old men, bustup angle, closeup,, ((analog photo)), (detailed), ZEISS, studio quality, 8k, 4k, uhd, (((photorealistic))), ((detailed)), transfer, dribble, precisionism, associated press photo, award-winning

Negative prompt [redacted] ((painting)), ((drawing)), ((sketch)), ((camera)), ((rendering)), mutilated, disfigured, unnatural, bad anatomy, morbid, too many fingers, deformed palm, deformed arm, deformed limbs, extra limbs, missing limbs, unnatural pose,(((cropped))), (((watermark))), ((logo)), ((barcode)), ((UI)), ((signature)), ((text)), ((label)), ((error)), ((title)), stickers, markings, speech bubbles, lines, cropped, lowres, low quality, artifacts



IV. 画像生成の手法 i2i 機能：インペイント

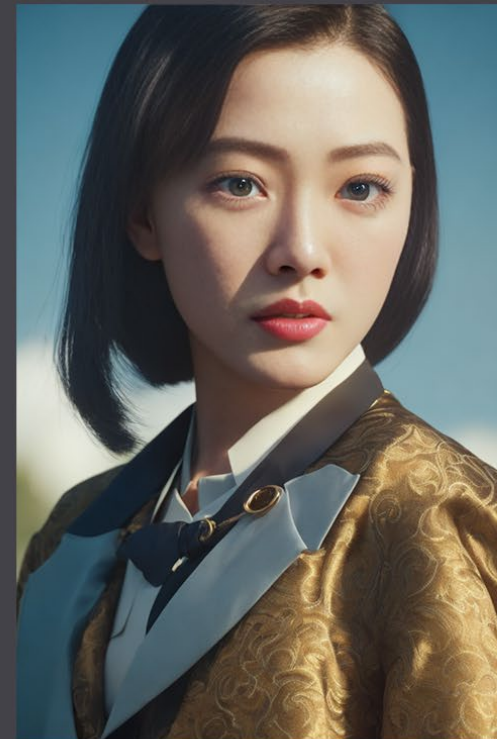
インペイントは欠けている部分や削除したい部分に対して、周囲の情報や画像全体の背景知識に基づいて適切なピクセル情報を生成し、画像を自然に補完する。画像生成時のモデルやプロンプト、シード値などを必ず合わせる必要は無く、元画との馴染みはAI側がある程度調整してくれる。

Prompt: A photograph of A photograph of best quality, ultra high res, (photorealistic:1.4), 1girl, kind eyes, < [REDACTED] >, ((analog photo)), (detailed), ZEISS, studio quality, 8k, 4k, uhd, (((photorealistic))), ((detailed)), transfer, ((analog photo)), (detailed), ZEISS, studio quality, 8k, 4k, uhd, (((photorealistic))), ((detailed)), transfer

[REDACTED], paintings, sketches, (worst quality:2), (low quality:2), (normal quality:2), lowres, normal quality, ((monochrome)), ((grayscale)), skin spots, acnes, skin blemishes, age spot, glans, ((painting)), ((drawing)), ((sketch)), ((camera)), ((rendering)), (((cropped))), (((watermark))), ((logo)), ((barcode)), ((UI)), ((signature)), ((text)), ((label)), ((error)), ((title)), stickers, markings, speech bubbles, lines, cropped, lowres, low quality, artifacts, ((painting)), ((drawing)), ((sketch)), ((camera)), ((rendering)), (((cropped))), (((watermark))), ((logo)), ((barcode)), ((UI)), ((signature)), ((text)), ((label)), ((error)), ((title)), stickers, markings, speech bubbles, lines, cropped, lowres, low quality, artifacts



Inpaint Mask



IV. 画像生成の手法 i2i 機能：アウトペイント

アウトペイントは画像を延長したい時に使う機能。画像生成時のモデルやプロンプト、シード値などを合わせることで延長部分の画像の精度は上がる。プロンプトやシード値が判らない（or無い）画像でも、fill in（埋める）などのシンプルなプロンプトで延長可能。

Prompt: A photograph of mdjrny-v4, 3d alpha of intricate steampunk designed car, detailed textures, 3d textures, ((analog photo)), (detailed), ZEISS, studio quality, 8k, 4k, uhd, (((photorealistic))), ((detailed)), transfer, dribbble, precisionism, associated press photo, award-winning

Negative prompt: , ((painting)), ((drawing)), ((sketch)), ((camera)), ((rendering)), mutilated, disfigured, unnatural, bad anatomy, morbid, too many fingers, deformed palm, deformed arm, deformed limbs, extra limbs, missing limbs, unnatural pose,(((cropped))), (((watermark))), ((logo)), ((barcode)), ((UI)), ((signature)), ((text)), ((label)), ((error)), ((title)), stickers, markings, speech bubbles, lines, cropped, lowres, low quality, artifacts



Prompt: fill in



IV. 画像生成の手法 ControlNet（レイアウト制御）

t2iで人物のポーズやカメラアングル、画角の中での位置やサイズなどの画面レイアウトをテキスト（プロンプト）だけで完全にコントロールすることは大変難しい。i2iでもプロンプトの効き具合を強くするとレイアウトが崩壊する。これらの課題を解決する手法がControlNetという機能。

Prompt: best quality, ultra high res, (photorealistic:1.4), 1girl knit, short hair, (closeup), wide angle, (low angle), (looking at viewer), < ██████████ >, no background,

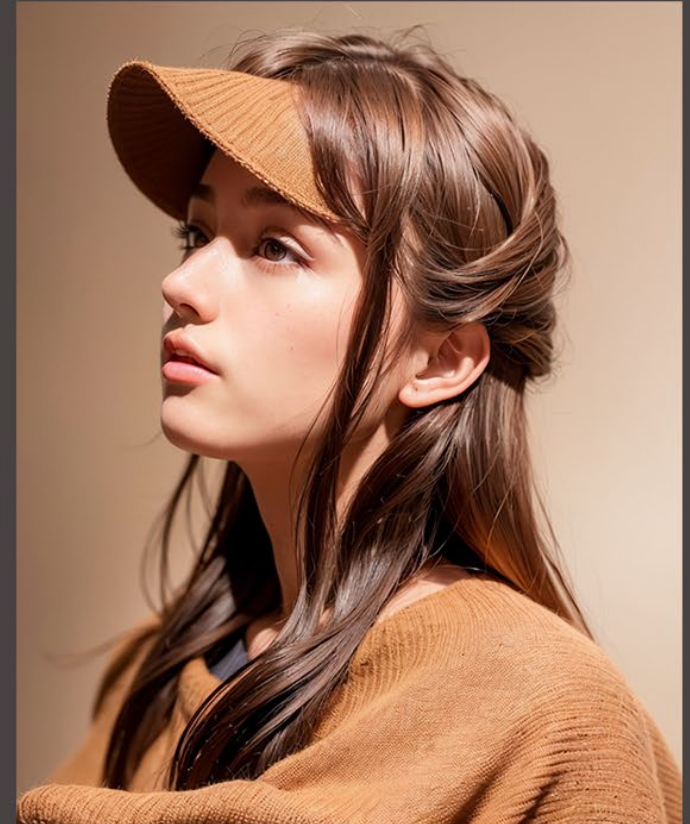
Negative prompt: ██████████, paintings, sketches, (worst quality:2), (low quality:2), (normal quality:2), lowres, normal quality, ((monochrome)), ((grayscale)), skin spots, acnes, skin blemishes, age spot, glans



IV. 画像生成の手法 ControlNet (レイアウト制御)

Prompt: best quality, ultra high res, (photorealistic:1.4), 1girl, on cap, short hair, (closeup), wide angle, (low angle), (looking at viewer), < [redacted] >, no background,

Negative prompt: [redacted], paintings, sketches, (worst quality:2), (low quality:2), (normal quality:2), lowres, normal quality, ((monochrome)), ((grayscale)), skin spots, acnes, skin blemishes, age spot, glans



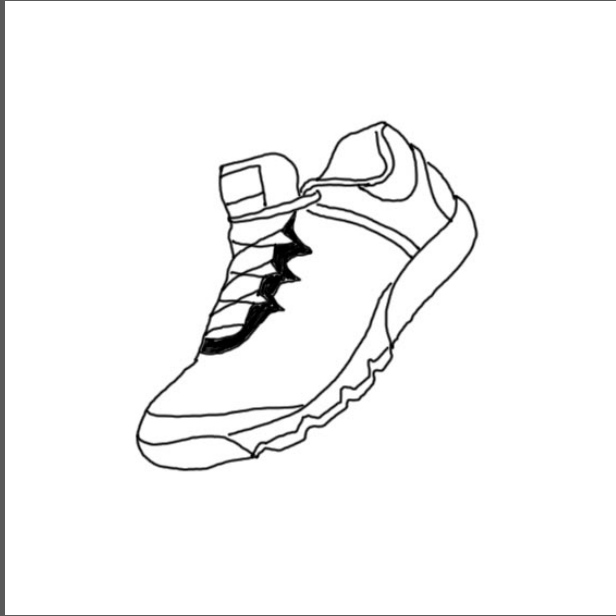
IV. 画像生成の手法 ControlNet (レイアウト制御)

Prompt: A photograph of a white dog is sniffing a vintage record player on a table top with a record player on it's side,steampunk,, ((analog photo)), (detailed), ZEISS, studio quality, 8k, 4k, uhd, (((photorealistic))), ((detailed)), transfer

Negative prompt: , ((painting)), ((drawing)), ((sketch)), ((camera)), ((rendering)),(((cropped))), (((watermark))), ((logo)), ((barcode)), ((UI)), ((signature)), ((text)), ((label)), ((error)), ((title)), stickers, markings, speech bubbles, lines, cropped, lowres, low quality, artifacts



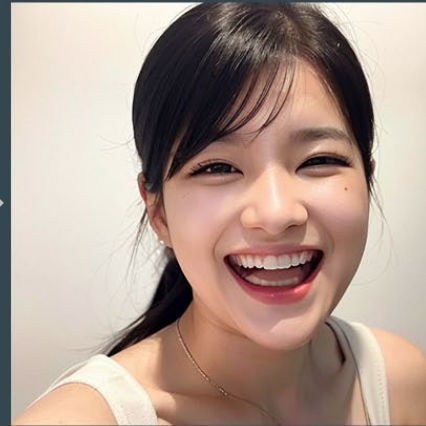
IV. 画像生成の手法 ControlNet (レイアウト制御)



IV. 画像生成の手法 ControlNet（表情制御）

ControlNetの機能を使って表情を制御することも出来る。

ControlNet : Face tracking



V. モデルへの追加学習

既存のAIモデルに任意の画像を追加学習させることが可能。教師データとなる画像は10枚程度が良い。



2023: a happy new year VONSpictures AI <https://www.vons.co.jp/>

Teacher Data



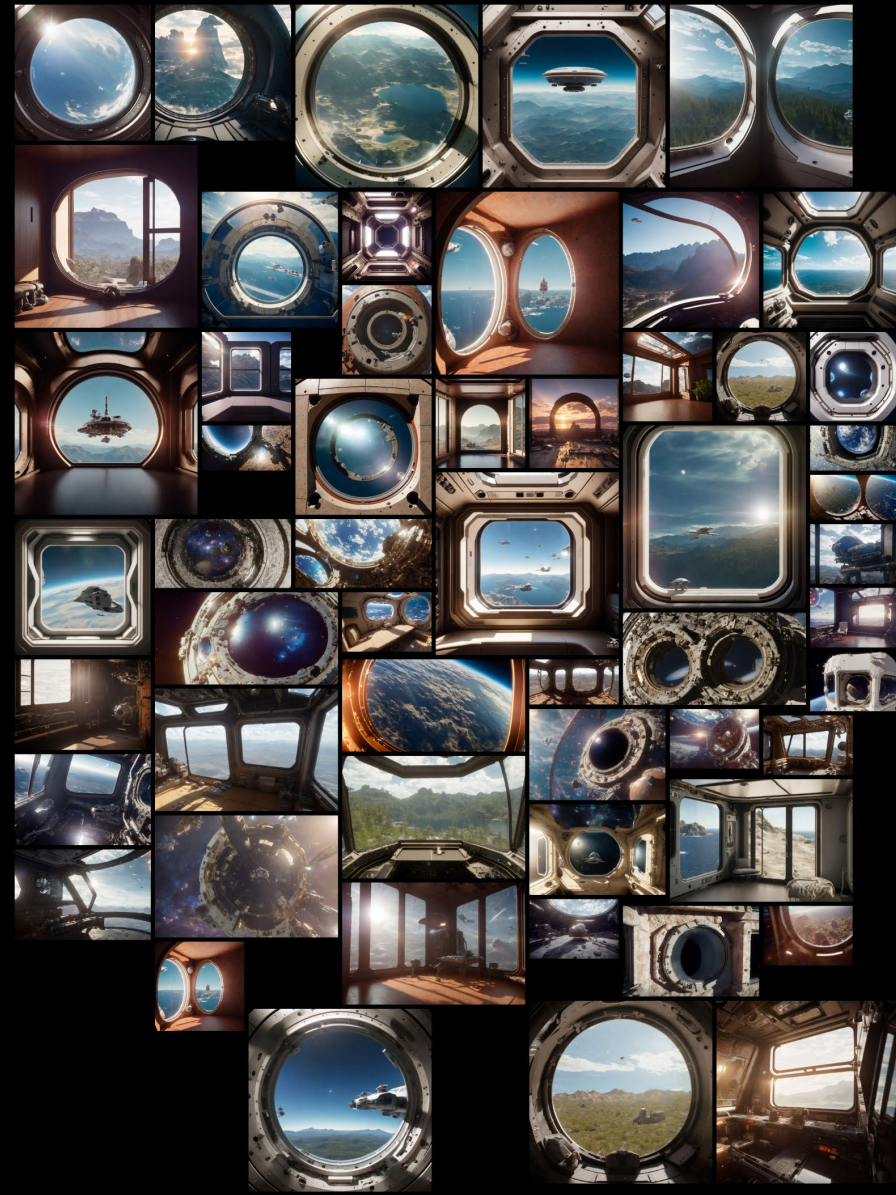
Prompt:

A photograph of redshift style, A dogvonta ((wearing a space suit helmet)), jumping on the moon, , sharp focus, ((full body)), ((analog photo)), (detailed), ZEISS, studio quality, 8k, 4k, uhd, (((photorealistic))), ((detailed)), transfer, dribbble, precisionism, associated press photo, award-winning Negative prompt: , ((painting)), ((drawing)), ((sketch)), ((camera)), ((rendering)), mutilated, disfigured, unnatural, bad anatomy, morbid, too many fingers, deformed palm, deformed arm, deformed limbs, extra limbs, missing limbs, unnatural pose,((cropped)), ((watermark)), ((logo)), ((barcode)), ((UI)), ((signature)), ((text)), ((label)), ((error)), ((title)), stickers, markings, speech bubbles, lines, cropped, lowres, low quality, artifacts



Prompt:

A photograph of redshift style, a beautiful intricate fine art, View of a spacecraft window from the outside of a spacecraft in outer space, centered, symmetrical, analog photo, film grain, , ((analog photo)), (detailed), ZEISS, studio quality, 8k, 4k, uhd, (((photorealistic))), ((detailed)), transfer Negative prompt: fog, back light, , ((painting)), ((drawing)), ((sketch)), ((camera)), ((rendering)), (((cropped))), (((watermark))), ((logo)), ((barcode)), ((UI)), ((signature)), ((text)), ((label)), ((error)), ((title)), stickers, markings, speech bubbles, lines, cropped, lowres, low quality, artifacts





2023: a happy new year VONSpictures AI <https://www.vons.co.jp/>

VII. 今後の展望

G7サミットに於ける「**広島AIプロセス**」の立上げ 2023年内に取りまとめる

岸田首相「**AI戦略会議**」で関係閣僚に早急な検討を指示 2023.5.11

経団連 AI活用戦略～**AI-Readyな社会の実現に向けて**～ 2019

新経済連盟（三木谷浩史・代表理事）は『**JX（Japan Transformation）宣言**～新経済連盟が考える日本の未来ビジョン～』を、岸田文雄・内閣総理大臣に提出 2023.7.14

経済産業省がさくらインターネットの生成AI向け**GPUクラウド事業**に68億円助成 2023.6.16

経済産業省がソフトバンクの**生成AIスパコン事業**に53億円助成 2023.7.07

日本画像生成AIコンソーシアム設立（JIGAC） 2023.6.20

VII. 今後の展望

画像生成AIの技術が急速に進化しているため、現行の法律や規制が追いついていないのが現状です。画像生成AIの利用に関する法的責任や規制が不明確であることで、法的な問題が発生する可能性があります。

一方で音楽業界ではCDやレコードからストリーミング配信に変化していく中でライセンス制度の発展やDigital Millennium Copyright Act (DMCA) の適用等により法的リスクが減少しました。

今後画像生成AIでも法律が整備され安全に社会実装される事が期待されます。

以上の事から最終成果物を画像生成AIだけで制作するのは時期尚早と判断します。しかし現状でも著作権侵害やプライバシー侵害に留意しながら画像生成AIの利用シーンは様々あります。

コンセプトビジュアル

広告 캠프制作

背景やパーツ

画像処理