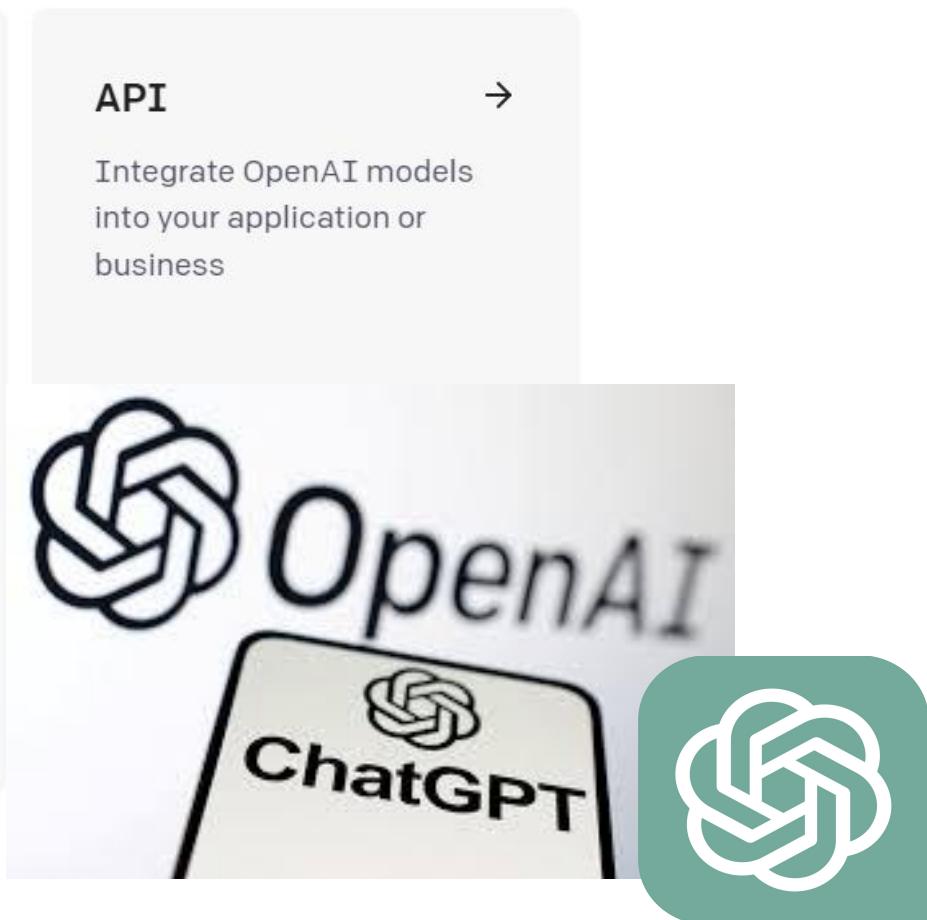
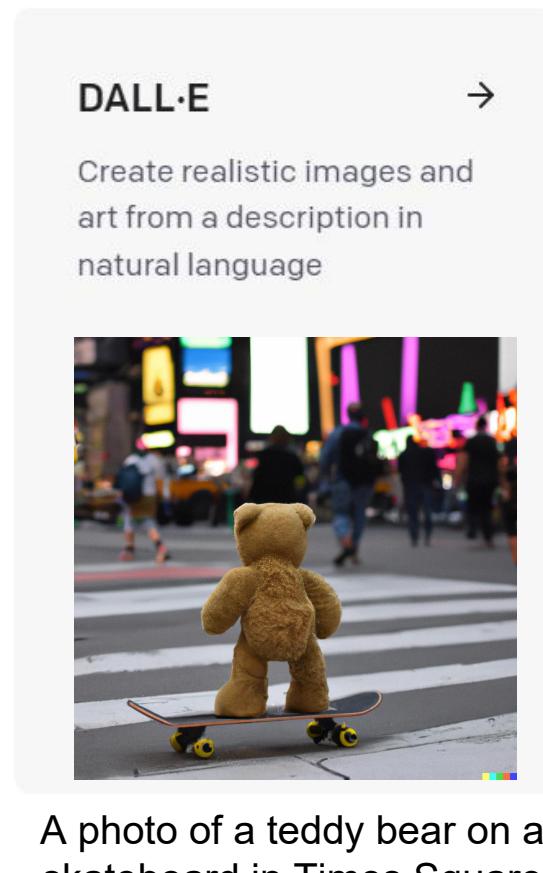
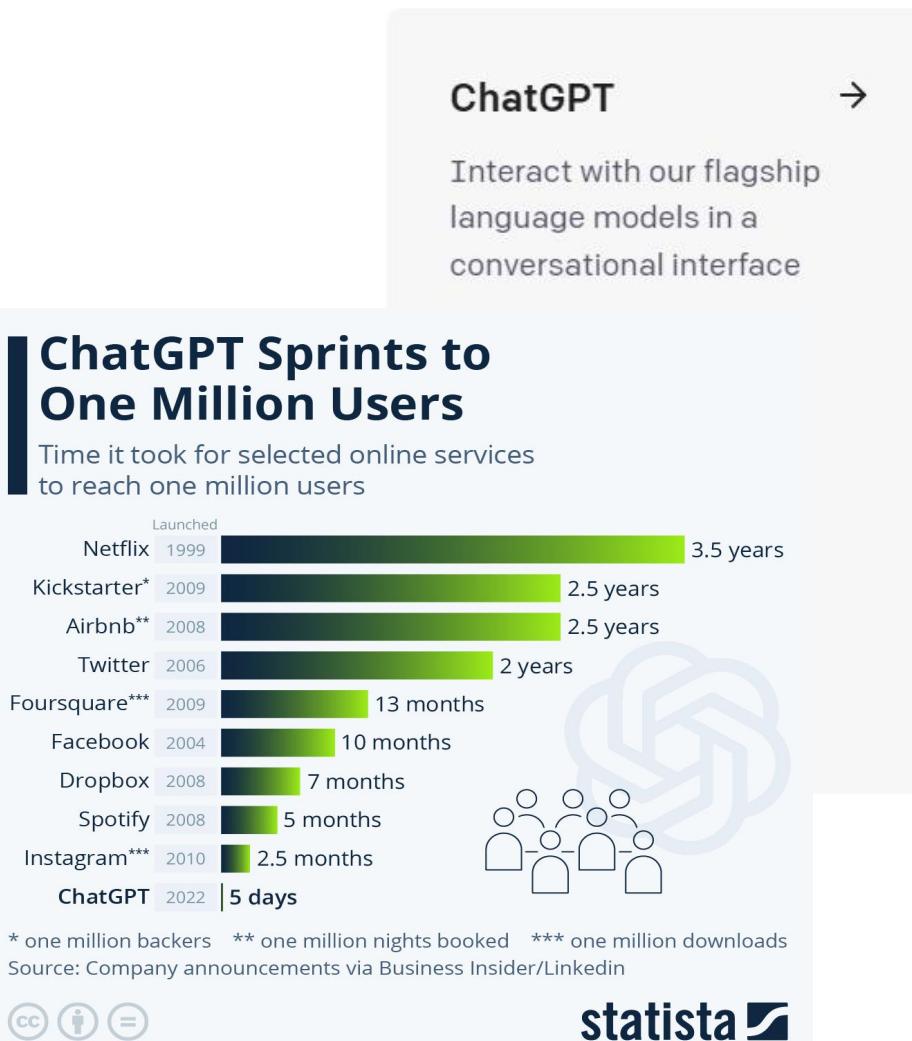


Generative AI Fine Tuning for Multimedia Designer

lbg@dongseo.ac.kr

ChatGPT

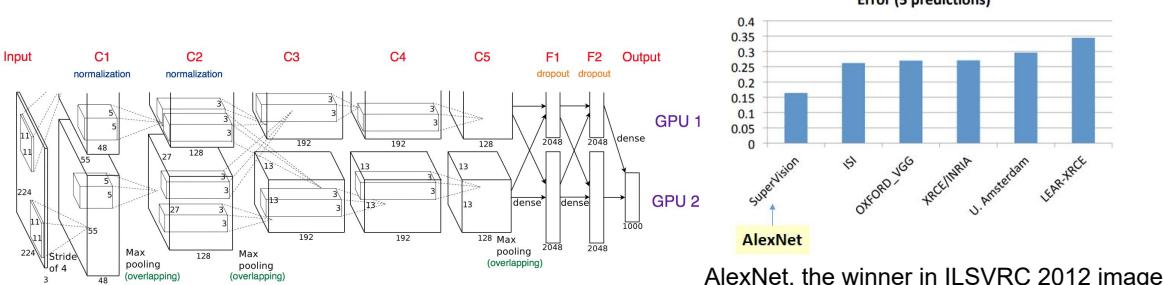
ChatGPT is an artificial intelligence (AI) chatbot developed by OpenAI.
ChatGPT gained one million users five days after launching in November 2022.





OpenAI is an American artificial intelligence (AI) research laboratory consisting of the non-profit **OpenAI Incorporated** and its for-profit subsidiary corporation **OpenAI Limited Partnership**. OpenAI conducts AI research with the declared intention of promoting and developing a friendly AI. OpenAI systems run on an Azure-based supercomputing platform from Microsoft.^{[6][7][8]}

OpenAI was founded in 2015 by Ilya Sutskever, Greg Brockman, Trevor Blackwell, Vicki Cheung, Andrej Karpathy, Durk Kingma, Jessica Livingston, John Schulman, Pamela Vagata, and Wojciech Zaremba, with Sam Altman and Elon Musk serving as the initial board members.^{[9][2][10]} Microsoft provided OpenAI LP with a \$1 billion investment in 2019 and a \$10 billion investment in 2023.^{[11][12]}

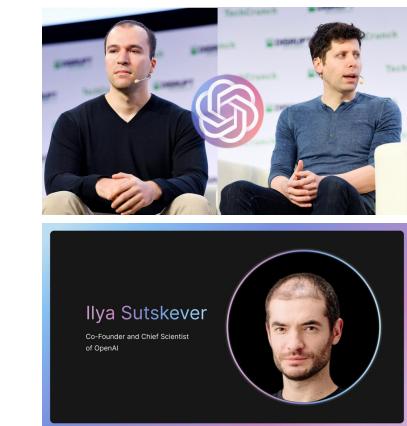


Key employees:

- CEO and co-founder:^[48] Sam Altman, former president of the startup accelerator Y Combinator
- President and co-founder:^[49] Greg Brockman, former CTO, 3rd employee of Stripe^[50]
- Chief Scientist and co-founder: Ilya Sutskever, a former Google expert on machine learning^[50]
- Chief Technology Officer:^[49] Mira Murati, previously at Leap Motion and Tesla, Inc.
- Chief Operating Officer:^[49] Brad Lightcap, previously at Y Combinator and JPMorgan Chase

Board of the OpenAI nonprofit:

- Greg Brockman
- Ilya Sutskever
- Sam Altman
- Adam D'Angelo
- Will Hurd
- Tasha McCauley
- Helen Toner
- Shivon Zilis



Individual investors:^[50]

- Reid Hoffman, LinkedIn co-founder^[51]
- Peter Thiel, PayPal co-founder^[51]
- Jessica Livingston, a founding partner of Y Combinator

Corporate investors:

- Microsoft^[52]
- Khosla Ventures^[53]
- Infosys^[54]



November 2023.

OpenAI fires co-founder and CEO Sam Altman for allegedly lying to company board

November 2023.

AI firm's board said Altman was 'not consistently candid in his communications with the board' and had lost its confidence

● [Interview: Sam Altman on whether AI will save humanity - or destroy it](#)



Sam Altman, CEO of OpenAI, attends the Asia-Pacific Economic Cooperation (Apec) CEO Summit in San Francisco, California. Photograph: Carlos Barría/Reuters

BREAKING / ARTIFICIAL INTELLIGENCE / TECH

OpenAI board in discussions with Sam Altman to return as CEO

Y Combinator (YC) is an American technology startup accelerator launched in March 2005.^[1] It has been used to launch more than 4,000 companies,^[2] including Airbnb, Coinbase, Cruise, DoorDash, Dropbox, Instacart, Quora, PagerDuty, Reddit, Stripe and Twitch.^[3] The combined valuation of the top YC companies was more than \$600 billion by January 2023.^[4] The company's accelerator program started in Boston and Mountain View, expanded to San Francisco in 2019, and was entirely online during the COVID-19 pandemic.^[5] Forbes characterized the company in 2012 as one of the most successful startup accelerators in Silicon Valley.^[6]

TECHNOLOGY | UPDATED 6:54 P.M.

Why Was Sam Altman Fired As CEO of OpenAI?

By Chas Danner, associate editor at Intelligencer

Sam Altman, the co-founder and CEO of OpenAI who many considered to be the leader of the burgeoning artificial-intelligence revolution, was fired on Friday. The dismissal followed what the company said was a “deliberative review process by the board, which concluded that he was not consistently candid in his communications with the board, hindering its ability to exercise its responsibilities.” As a result, “the board no longer has confidence in his ability to continue leading OpenAI.” Soon after, OpenAI president and Altman-ally Greg Brockman — who the company also removed as chairman of the board — resigned, as did three senior OpenAI researchers.

TECHNOLOGY

OpenAI Investors Trying to Get Sam Altman Back as CEO After Sudden Firing

The co-founder of the artificial-intelligence company behind ChatGPT was pushed out by the board Friday

By Keach Hagey [Follow](#), Berber Jin [Follow](#) and Deepa Seetharaman [Follow](#)

Updated Nov. 18, 2023 7:58 pm ET



Ilya Sutskever
He is a world-famous researcher in the field of artificial intelligence.

Superalignment is the process of supervising, controlling and governing artificial superintelligence systems. Aligning advanced AI systems with human values and goals can help prevent them from exhibiting harmful and uncontrollable behavior.

일리야 수츠케버(37세)

| 출생지 | 국적 | 학력 |
|---|-------------------|--|
| -구소련 니즈니노브고로드 -다섯 살 때 가족이 이스라엘 예루살렘으로이주 | 캐나다·이스라엘 이중 국적 | -이스라엘 오픈대학교 -토론토대 수학 학사 -토론토대 컴퓨터 과학 석·박사 -스탠퍼드대 박사 후 연구원 |

주요 이력

- 제프리 힌턴 토론토대 교수와 딥러닝의 시초 '알렉스넷' 개발
- 구글 딥마인드에서 기계 학습 엔진 '텐서 플로'·바둑 AI '알파고' 개발 주도
- 오픈AI에서 AI를 제어하는 '수퍼 얼라인먼트' 팀 구축

"The future will undoubtedly be more AI-friendly. However, no one knows whether it will be friendly to humanity as well."

"미래는 분명 AI 친화적으로 발전할 것이다. 하지만 그것이 인류에게도 친화적일지는 아무도 모른다."

"인간의 기능을 더 뛰어나게 수행하는 일반 인공지능(AGI)이 가까운 시일 내 출현할 가능성은 매우 크다."

"AGI의 신념과 욕망이 인류가 바라는 바와 같지 않게 된다면 재앙이다."

"It would be a disaster if the beliefs and desires of AGIs were not the same as those of humanity."

Safe Superintelligence Inc.

| | |
|----------------------------|---|
| Industry | Artificial Intelligence |
| Founded | June 19, 2024 |
| Founders | Ilya Sutskever Daniel Gross Daniel Levy |
| Headquarters | Palo Alto, California Tel Aviv, Israel |
| Number of employees | 20 (March 2025) |
| Website | ssi.inc |

[Sequence to Sequence Learning with Neural Networks](#), Ilya Sutskever, Oriol Vinyals, Quoc V. Le, NIPS 2014 Cited by 29579

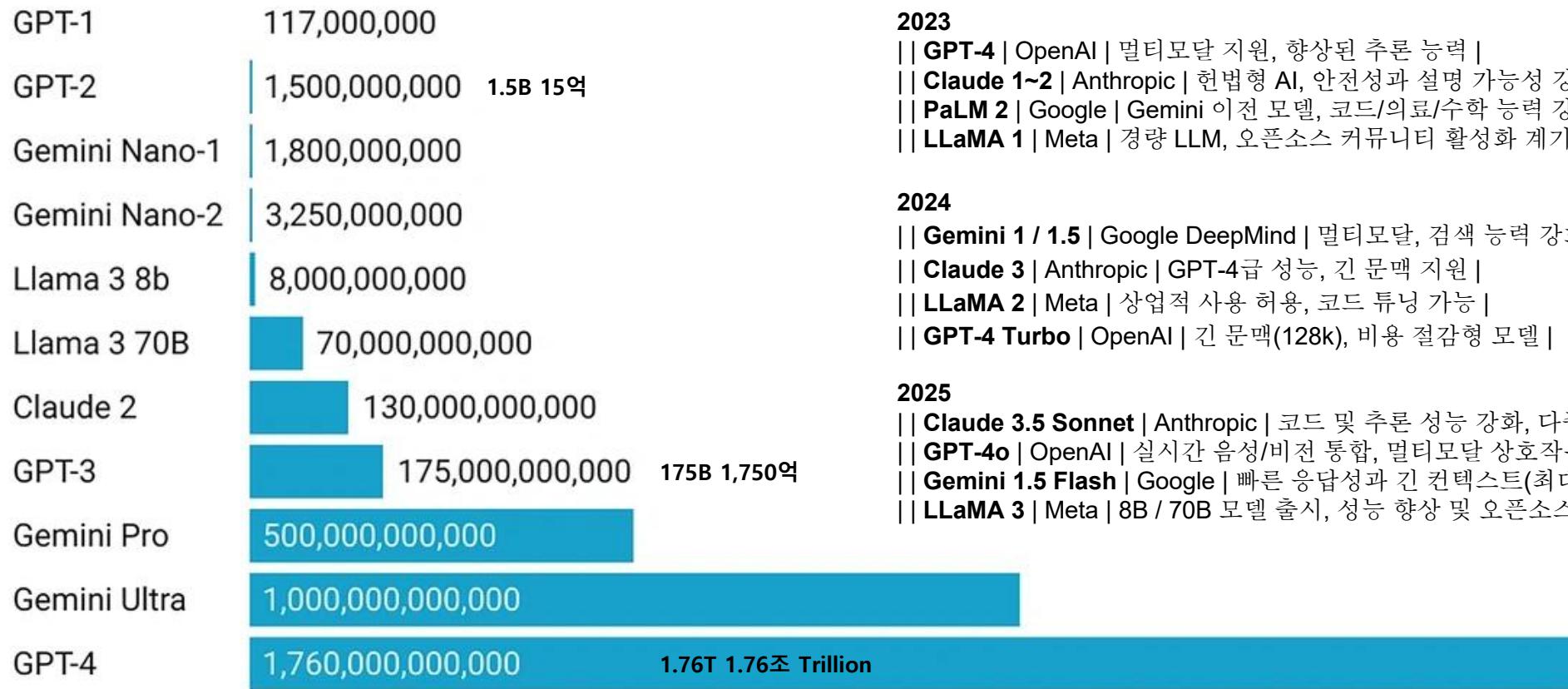
[Mastering the game of Go with deep neural networks and tree search](#). Silver, D., Huang, A., Ilya Sutskever. et al. Nature 529, 484–489 (2016).Cited by 22407

[Learning Transferable Visual Models From Natural Language Supervision](#), Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, Ilya Sutskever, CVPR2021 Cited by 46293

In 2016, Google introduced GNMT (Google Neural Machine Translation) and switched to a Seq2Seq-based neural network model.

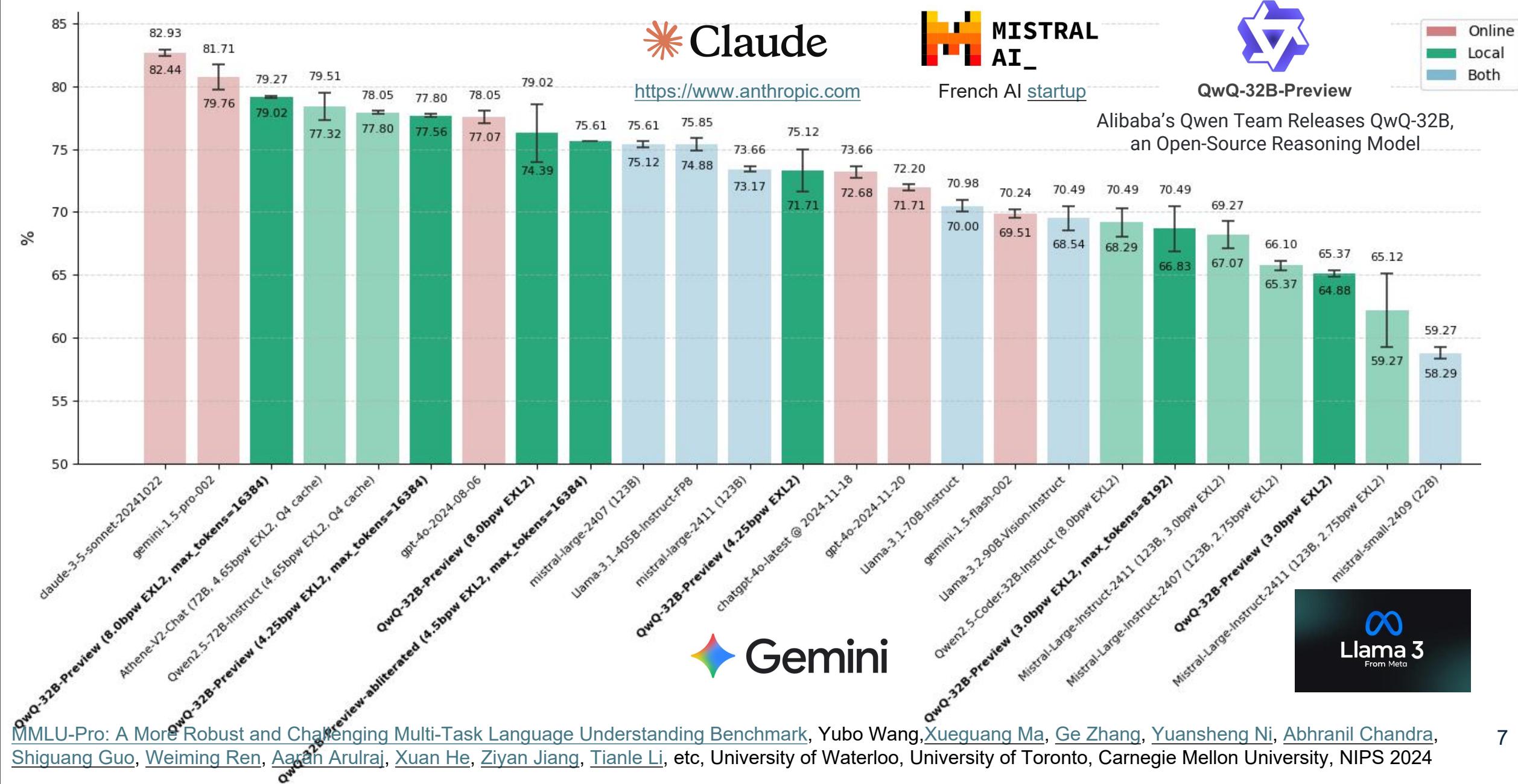
Parameters in Selected AI Models

Some of these figures are estimates. Newer models are many times larger than their predecessors.



LLM performance is advancing to a level similar to humans in terms of language comprehension, generation, reasoning, and problem-solving abilities, and even surpassing humans in some tasks. However, improving accuracy, stability, and reliability remain key research challenges.

Wolfram Ravenwolf's MMLU-Pro Computer Science LLM Benchmark Results (2024-12-04)



This LLM leaderboard displays the latest public benchmark performance for SOTA model versions released after April 2024. The data comes from model providers as well as independently run evaluations by Vellum or the open source community. We feature results from non-saturated benchmarks, excluding outdated benchmarks (e.g. MMLU).

See how leading models stack up across text, image, vision, and beyond. This page gives you a snapshot of each Arena, you can explore deeper insights in their dedicated tabs.

| Models | GRIND | AIME 2024 | GPQA | SWE Bench | MATH 500 | BFCL | Alder Polyglot |
|--|-------|-----------|-------|-----------|----------|------|----------------|
|  Grok 4 | n/a | 94% | 87.5% | 75% | n/a | n/a | n/a |
|  Claude 4 Opus | 67.9% | n/a | 79.6% | 72.5% | n/a | n/a | n/a |
|  Claude 4 Sonnet | 75% | n/a | 75.4% | 72.7% | n/a | n/a | n/a |
|  Gemini 2.5 Flash | n/a | 88% | 78.3% | n/a | n/a | n/a | 511% |
|  OpenAI o3 | n/a | 91.6% | 83.3% | 691% | n/a | n/a | 81.3% |
|  OpenAI o4-mini | 50% | 93.4% | 81.4% | 68.1% | n/a | n/a | 68.9% |
|  Nemotron Ultra 253B | 57.1% | 80.08% | 76% | n/a | n/a | n/a | n/a |
|  GPT-4.1 nano | n/a | 29.4% | 50.3% | n/a | n/a | n/a | 9.8% |
|  GPT-4.1 mini | n/a | 49.6% | 65% | 23.6% | n/a | n/a | 34.7% |
|  GPT-4.1 | n/a | 48.1% | 66.3% | 55% | n/a | n/a | n/a |
|  Llama 4 Behemoth | n/a | n/a | 73.7% | n/a | 95% | n/a | n/a |

| Rank (UB) ↑ | Model ↑ | Score ↑ | Votes ↑ |
|--------------------------|---|---------|---------|
| 1 |  gemini-2.5-pro | 1458 | 25,480 |
| 1 |  o3-2025-04-16 | 1451 | 31,450 |
| 2 |  chatgpt-4o-latest-20250326 | 1442 | 30,344 |
| 3 |  gpt-4.5-preview-2025-02-27 | 1438 | 15,271 |
| 3 |  qwen3-235b-a22b-instruct-2507 | 1433 | 3,386 |
| 4 |  grok-4-0709 | 1429 | 12,591 |
| 5 |  claude-opus-4-20250514-think... | 1420 | 17,512 |
| 5 |  kimi-k2-0711-preview | 1420 | 10,934 |
| 5 |  deepseek-r1-0528 | 1417 | 17,934 |
| 7 |  claude-opus-4-20250514 | 1412 | 25,729 |
| View all | | | |

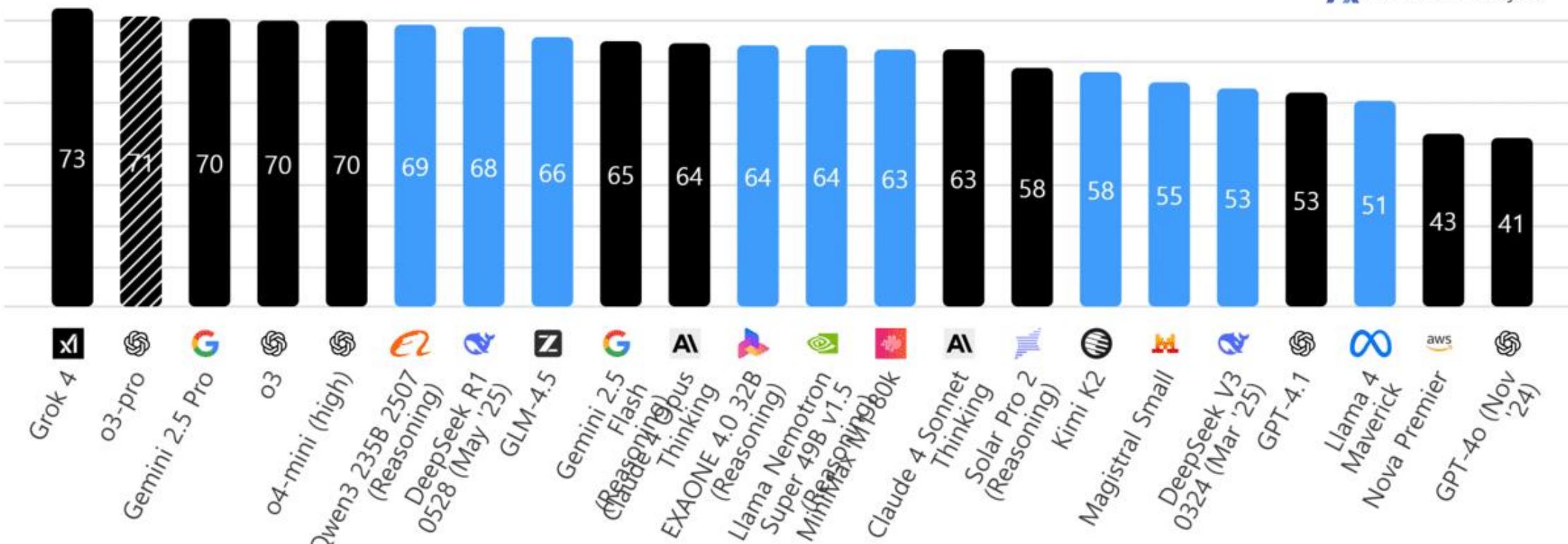
Artificial Analysis Intelligence Index by Open Weights vs Proprietary

Artificial Analysis Intelligence Index incorporates 7 evaluations: MMLU-Pro, GPQA Diamond, Humanity's Last Exam, LiveCodeBench, SciCode, AIME, MATH-500

■ Estimate (independent evaluation forthcoming)

■ Proprietary ■ Open Weights

Artificial Analysis



2차 평가 종료, 국가대표 AI 최종 승부 시작... 10개 캔소시엄, 8월 초 5곳으로 압축

2025년 7월, 대한민국 인공지능(AI) 생태계는 거대한 도약을 준비 중이다. 과학기술정보통신부가 주관하는 ‘국가대표 AI 파운데이션 모델’ 개발 공모에 국내 유수의 AI 기업과 연구기관들이 참여한 것이다. 총 15개 캔소시엄이 제안서를 제출하며 예상을 뛰어넘는 뜨거운 경쟁을 벌였다. 이는 단순히 기술적 진보를 넘어, 디지털 주권 확보와 글로벌 AI 경쟁력 강화를 위한 전략적 포석으로 읽힌다. 이번 사업은 “글로벌 프론티어 모델 성능의 95% 이상”이라는 도전적 목표를 내걸며, 대한민국이 AI 주권국가로 발돋움할 수 있는 분기점이 될 것이다.

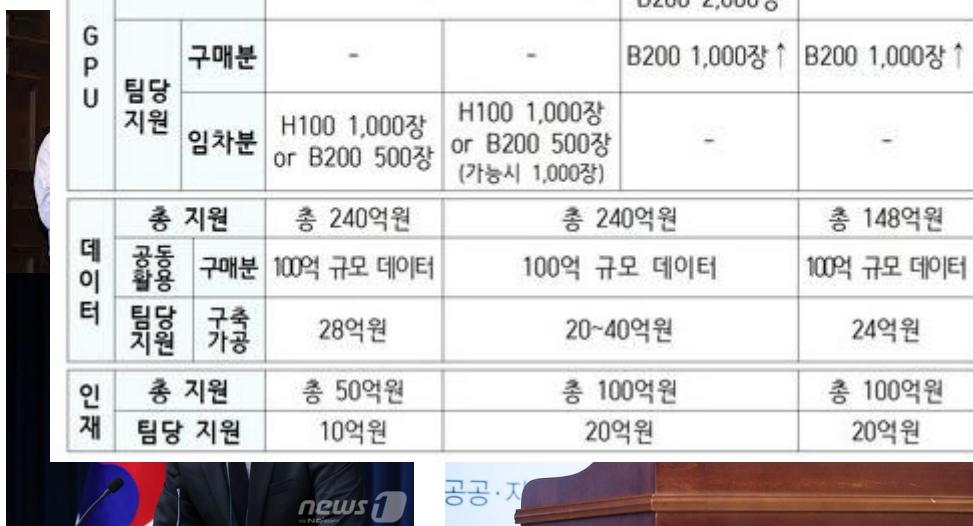
| 지원규모 (안) | | 1차 ('25.下) | 2차 ('26.上) | 3차 ('26.下) | 4차 ('27년) |
|-------------------------|----------------|-----------------------------|---|----------------------------|---------------|
| G P U 팀당 지원 | 총 지원 | H100 1,000장, B200 1,000장 | | H200 1,000장 B200 2,000장 | B200 2,000장 |
| | 구매분 | - | - | B200 1,000장 ↑ | B200 1,000장 ↑ |
| | 임차분 | H100 1,000장 or B200 500장 | H100 1,000장 or B200 500장 (가능시 1,000장) | - | - |
| 총 지원 | | 총 240억원 | 총 240억원 | 총 148억원 | |
| 데 이 터 공동 활용 | 구매분 | 100억 규모 데이터 | 100억 규모 데이터 | 100억 규모 데이터 | |
| | 팀당 지원 가공 | 28억원 | 20~40억원 | 24억원 | |
| 인 재 | 총 지원 | 총 50억원 | 총 100억원 | 총 100억원 | |
| | 팀당 지원 | 10억원 | 20억원 | 20억원 | |



This is seen as a strategic move to secure digital sovereignty and strengthen global AI competitiveness. This project has a challenging goal of achieving “more than 95% of global frontier model performance.”

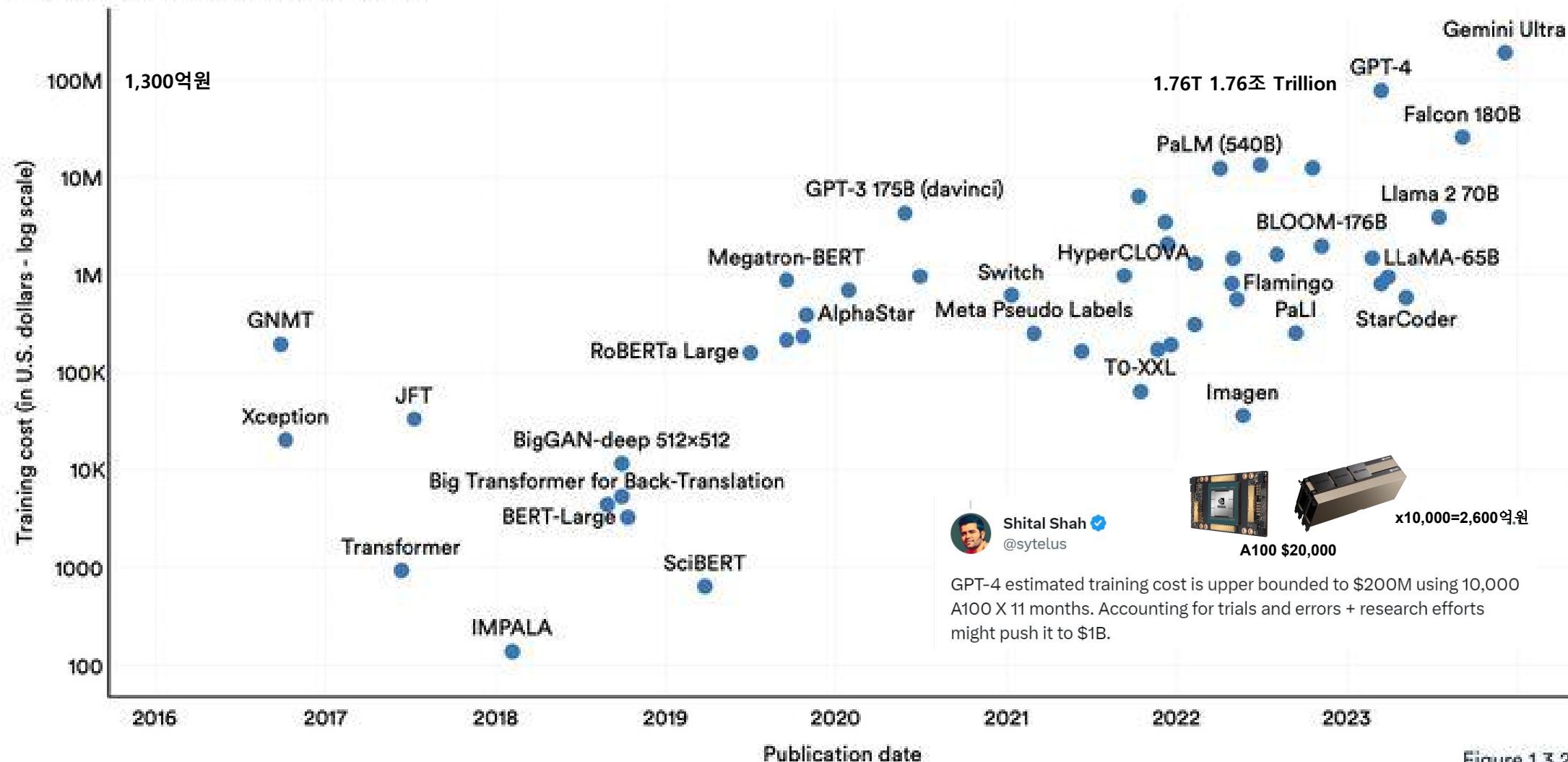
| 참여 기업 | 참여 기관 · 대학(연구팀) |
|-----------|--|
| 네이버클라우드 | 트웰브랩스 등 서울대, 포항공대, 고려대, 한양대 등 |
| 모티프테크놀로지스 | 모레, 삼일회계법인 등 서울대, KAIST 등 |
| 업스테이지 | 플리토, 래블업, 노타 등 서강대, KAIST 등 |
| SKT | 크래프톤, 포티투닷, 리밸리온, 라이너, 셀렉트스타, 서울대학교, 한국과학기술원 |
| NC AI | 고려대학교, 서울대학교, 연세대학교, 한국과학기술원, 한국전자통신연구원, 에이아이웍스, 포스코DX, 롯데이노베이트, HL로보틱스, 인터엑스, 미디어젠, 문화방송, NHN |
| LG AI연구원 | LG 유플러스, LG CNS, 슈퍼브AI, 퓨리오사AI, 프렌들리AI, 이스트소프트, 이스트에이드, 한글과컴퓨터, 뤼튼테크놀로지스, 미국 위스콘신대 메디슨 등 |

| 컨소시엄 주관사(가나다순) | 컨소시엄 참여기관 구성 |
|----------------|--|
| 네이버클라우드 | 네이버, 트웰브랩스, 서울대학교, 한국과학기술원, 포항공과대학교, 고려대학교, 한양대학교 |
| 업스테이지 | 노타, 래블업, 플리토, 뷰노, 마키나락스, 로앤컴퍼니, 오케스트로, 데이원컴퍼니, 올거나이즈코리아, 금융결제원, 서강대학교, 한국과학기술원 |
| SKT | 크래프톤, 포티투닷, 리밸리온, 라이너, 셀렉트스타, 서울대학교, 한국과학기술원 |
| NC AI | 고려대학교, 서울대학교, 연세대학교, 한국과학기술원, 한국전자통신연구원, 에이아이웍스, 포스코DX, 롯데이노베이트, HL로보틱스, 인터엑스, 미디어젠, 문화방송, NHN |
| LG AI연구원 | LG 유플러스, LG CNS, 슈퍼브AI, 퓨리오사AI, 프렌들리AI, 이스트소프트, 이스트에이드, 한글과컴퓨터, 뤼튼테크놀로지스, 미국 위스콘신대 메디슨 등 |



Estimated training cost of select AI models, 2016–23

Source: Epoch, 2023 | Chart: 2024 AI Index report



ChatGPT 4o

April 2025.

Generative Pre-trained Transformer 4 Omni (GPT-4o)

| | |
|-----------------|--|
| Developer(s) | OpenAI |
| Initial release | May 13, 2024; 10 months ago |
| Preview release | ChatGPT-4o-latest (2025-03-26) / March 26, 2025; 17 days ago |
| Predecessor | GPT-4 Turbo |
| Successor | OpenAI o1 GPT-4.5 |
| Type | Multimodal Large language model Generative pre-trained transformer Foundation model |
| License | Proprietary |
| Website | openai.com/index/hello-gpt-4o |



Examples of images generated by GPT-4o. These include a young girl at a farmer's market (left), and Karl Marx in a modern-day context (right).



← 게시물



Sam Altman @sama

it's super fun seeing people love images in chatgpt.

but our GPUs are melting.

we are going to temporarily introduce some rate limits while we work on making it more efficient. hopefully won't be long!

chatgpt free tier will get 3 generations per day soon.

IT·과학

"내 프사 지브리풍으로"...

ChatGPT 가입자 시간당 100만명 늘어

'미친 확산' 넘어선 ChatGPT...지브리 그림 한 장에 수익 30% 뛰어[지브리 모먼트]



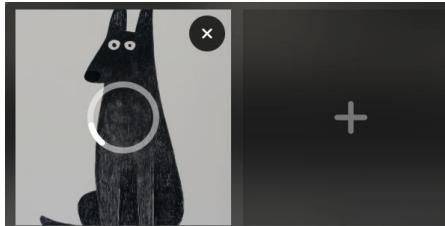


ChatGPT 4o





ChatGPT 4o



1컷:

(지친 표정으로 알람을 끄며 중얼거리는 주인공)
"아... 또 출근이야? 꼭 회사에 가야 할까?"

2컷:

(노트북으로 집에서 일하는 여유로운 모습)
"재택근무면 침대에서 바로 출근 가능인데..."

3컷:

(카페에서 노트북으로 일하며 커피 한 잔을 마시는 모습)
"카페 출근이면 기분 전환도 되잖아!"

4컷:

(다시 회사에서 쓸쓸하게 일하는 모습)
"하지만 현실은... 출근 안 하면 월급도 안 온다..."
(한숨 쉬며 생각하는 주인공)
"언젠간 나도 자유롭게 일할 수 있을까?"

위 내용을 담은 인스타툰을 그려줘, 다만 말풍선을 비워줘! 왜냐하면 내가 직접 각색해서 넣을 예정이거든!

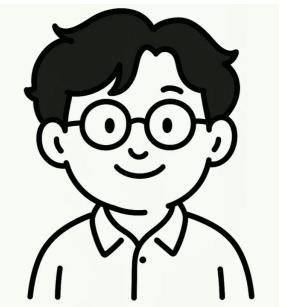




ChatGPT 4o



Please create a photo of the man in the first photo wearing the clothes in the second photo.



Draw this person with simple lines and cute characters.



ChatGPT 4o



Draw a set of emojis like the attached image, using the character above as the main character.

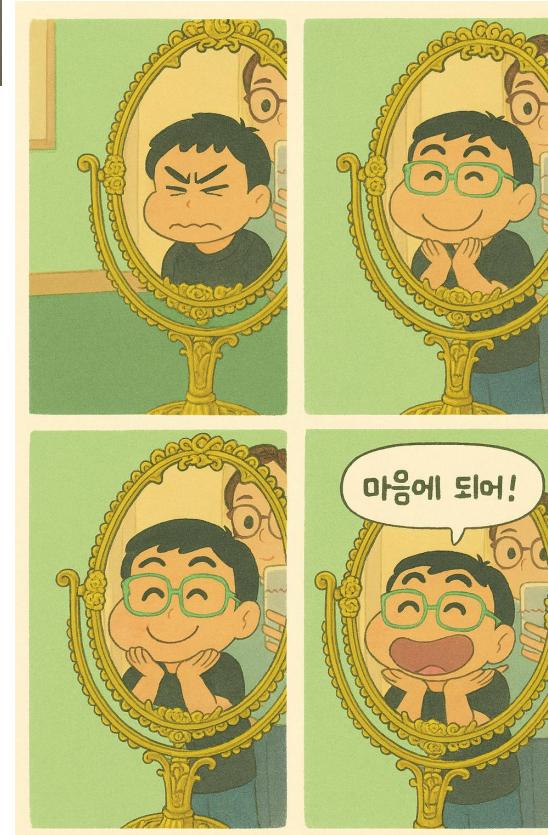


이미지를 짱구는 못말려 애니메이션의 스타일로 변경해줘
이미지에서 거울속에 비치는 아이를 짱구 그림체 스타일로 바꿔줘

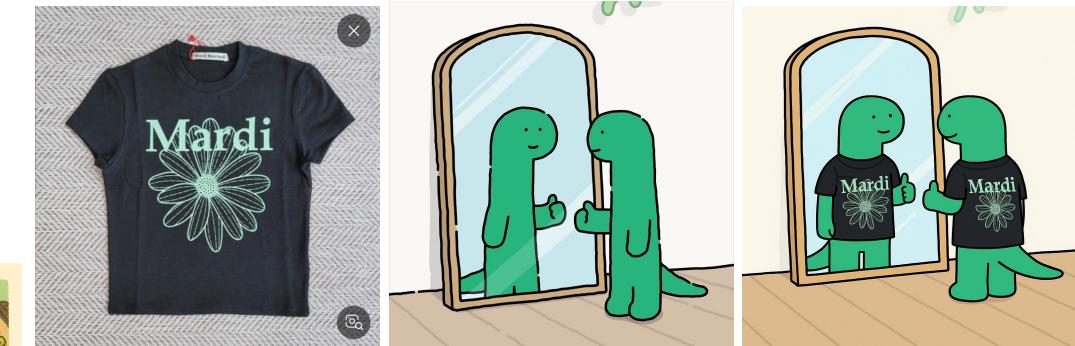


이 이미지를 참고로 해서
네컷 만화를 만들고 싶어.
장면 별 시나리오는

- [1. 안경을 안 쓰고 인상을 찡그린 표정]
2. 안경을 쓴 모습
3. 안경을 쓴게 만족스러운 표정 꽃받침 포즈
4. 말풍선에 '마음에 들어!'라는 글과 만세를 하는 포즈]
이렇게 될거야. 어린이들이 보는 동화느낌이 나면 좋겠어



캐릭터가 그려져 있는 사진에 티셔츠 사진을 합성해서 캐릭터가 티셔츠를 입고 있는 이미지를 만들어줘



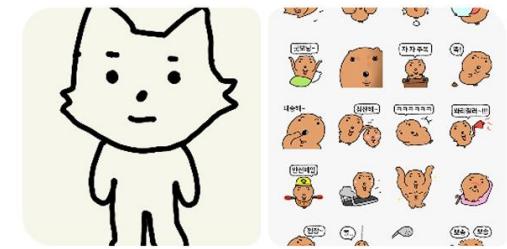
이 캐릭터로 카카오톡 이모티콘을 만들고 싶어. 캐릭터의 헤어나 얼굴 크기등의 특징을 통일해서 동일 캐릭터로 인지 할 수 있게 해줘. 10가지의 다양한 표정을 넣어서 상황에 맞게 사용 할 수 있는 이모티콘을 만들어줘.





- 1컷: (알람소리에 일어나는 주인공)으어어어
2컷: (폰으로 시간을 보는 모습)아 5분만 더 자고싶다,,
3컷: (침대에서 일어나는 주인공)아 자면 안돼
4컷: (화장실에서 세수하는 모습)어푸어푸

이미지 생성됨



Draw a set of emojis like the attached image, using the character above as the main character.

이미지 생성됨



Runway Gen-4, Pika 1.0



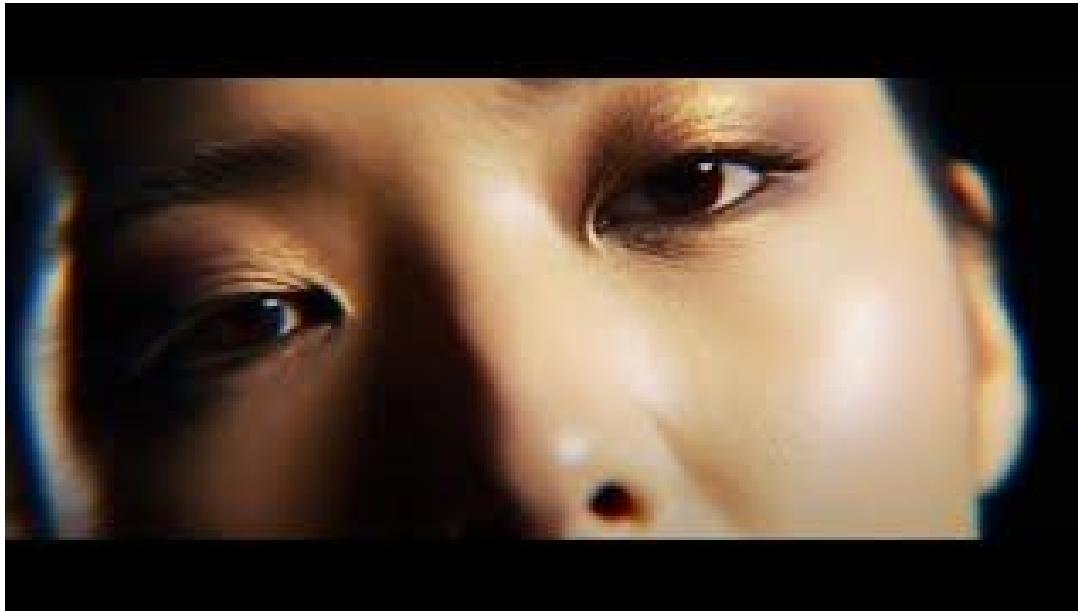
In February 2023 Runway released Gen-1 and Gen-2 is one of the first commercially available [text-to-video models](#). **Gen-4** On March 31, 2025, Runway released its latest flagship model, Gen-4.



PIKA LABS – PIKA 1.0 The founding team behind PicLab boasts impressive credentials. First, co-founder Demi Guo (郭文景) is a 95er (九五后 born after 1995) from Hangzhou, China (nationality: US). She was the first student from Zhejiang Province, to be pre-admitted to Harvard University in the US, and was featured on China's state-run media CCTV, earning her the title of “genius girl.”



Google Veo2, Pika 2.2



A spec ad for [@Sephora](#)'s imaginary new series,
PURE EDGE, created as part of a Google Veo2 test.



Transform your photos and videos with Pika 2.2—
introducing Pikaframes, and generations up to 10s.

The Silmarillion - Teaser Trailer



An AI generated Silmarillion trailer I put together using Midjourney, Luma and Kling.
Sound effects and voice overs in Eleven Labs. Music by Audiomachine.

고려대 세종캠퍼스, AI 영화 ‘걸리버 율도국 여행기’ 제6회 창원국제민주영화제 개막작 선정

작성일 2024.11.15 | 작성자 고려대학교 세종캠퍼스 | 조회수 272

고려대학교 세종캠퍼스(부총장 김영)가 AI 영화 '걸리버 율도국 여행기'가 16일, 제6회 창원국제민주영화제의 개막작으로 선정되어 상영된다. 대학가 내에서 AI 영화를 제작하여, 국제 영화제에 초청을 받아 개막작으로 선정된 것은 처음 있는 일이다.



Korea University Sejong Campus' AI film "Gulliver's Travels to Yuldo" selected as the opening film for the 6th Changwon International Democratic Film Festival

Script writing

- ChatGPT, Claude AI, etc.

Background Creation

- Runway Gen-3 Alpha, MidJourney, etc.

Character design and animation

- MidJourney, ChatGPT, etc.

Explosion and battle scenes

- Runway, Mid.Journey, etc.

Music and sound

- Suno AI, Eleven Labs, etc.

Post-production

- DaVinci Resolve, Adobe, etc.



△(왼쪽부터) 박진호 연구교수, AIMZ미디어 소휘수 감독(반도체물리전공 20),
이윤진 교수(글로벌학부), 조은솔·홍현·김주현 학생(응용물리학과)

이 작품은 영국 작가 조나단 스위프트의 '걸리버 여행기'와 조선 시대 허균이 창조한 이상국가 '율도국'을 결합하여 동서양의 이상 사회를 AI 기술을 통해 새롭게 재해석한 영화다. 이는 기존의 AI 영화들이 주로 공상과학이나 디스토피아 장르에 치우친 것과 달리, 동서양의 유토피아적 상상력을 결합한 독특한 역사극 형식으로 새로운 스토리텔링의 가능성을 제시한다. 또한 AI 기술이 인간의 보편적 가치를 실현하는 도구로서 갈등과 차별을 초월한 유토피아적 상상력을 제시하며 관객들에게 큰 울림을 선사할 것으로 기대된다.

이 영화는 AI 영화라는 점과 더불어, 제작에 참여한 모든 인원이 고려대학교 구성원이라는 점에서 더욱 의미가 있다. 4단계 BK21 ARMR 시스템·콘텐츠 융합연구단 소속 박진호 연구교수의 지휘 아래 응용물리학과 김주현, 홍현, 조은솔 학생이 시나리오를 집필했으며 고려대 출신 소휘수 감독이 대표로 있는 제작사 AIMZ미디어에서 AI 생성 프로그램을 활용해 각 장면을 시각적으로 완성했다. 고려대 글로벌학부 이윤진 교수 또한 핵심 철학인 '율도이즘(문화평등사상)'을 도출하는데 도움을 주었다.



https://www.instagram.com/heterotopia_of_ingong/ 이미지 생성 midjourney 영상생성 luma 편집 capcut

서울벤처대학원대학교 박은지 AI문화경영연구소 소장/주임교수 프랑스 AI 정상회의 Korean AI Artist 展 총괄 큐레이터 FAHIS in GangNam 예술총감



서울벤처대학원대학교
SEOUL VENTURE UNIVERSITY

AI Catverse



AI Catverse
Digital creator
AI Imaginational Super Art of Super Cats!
Welcome to AI Catverse,
where the magic of artif...

YouTube · AI Catverse
320K+ followers

AI Catverse

AI Catverse @aicatverse 320K subscribers • 135 videos AI Imaginational Super Art of Super Cats! more more instagram.com/aicatverse

The grid displays 16 video thumbnails from the AI Catverse YouTube channel, each with a title, view count, and timestamp:

- Kitten's McQueen Transformed into PLANE EATER MONSTER! #aicat #catlovers... 1.34:54
- School Kitten's Train Transformed into HEAD EATER MONSTER! #aicat #catlovers... 1.33:08
- POOR KITTEN BULLIED AT SCHOOL! Become Rich & Gets Revenge! #aicat... 1.34:49
- CUTE KITTEN STUCK IN A WASHING MACHINE! in California #aicat #catlovers... 1.33:13
- Dad Cat BETRAYED his Cat Family for LOVER! in California #aicat #catlovers #cute... 1.35:10
- Cute Kitten Left Under Hot Sun In Car! in California #aicat #catlovers #cute... 1.33:51
- Cute Kitten JOYRIDE MISHAP in California! #aicat #catlovers #cute... 1.32:42
- Shocking Powers: GRANNY & KITTEN'S Electric Accident! in Georgia #cat... 1.33:26
- Brother Kitten took Revenge for her Sister! in Miami #aicat #catlovers #cute... 1.32:02
- Two Adorable Kittens Unfairly Abused By Cat Teacher! #aicat #catlovers #catcute... 1.35:40
- Kitten Transformed into a FROG MONSTER After Eat Lollipop! #aicat #catlovers... 1.34:26
- Police Cat Family Died in House! in Los Angeles #aicat #catlovers #cute... 1.33:11

Instagram



aicatverse [Follow](#) [Message](#) ...

119 posts 65.8K followers 0 following

AI Catverse
Digital creator
Ai Imaginational Super Art of Super Cats! 🐱
linktr.ee/aicatverse

<https://openai.com/ko-KR/sora/> <https://www.youtube.com/watch?v=Qt1Qw2GChgg&list=LL>

http://drive.google.com/file/d/1C3D5IK3qmQ0_uQqazgAS3olpb_3DHX-Y/view

http://drive.google.com/file/d/1CIMj_jAoo9dl5pH9dAMfNye4xmk8job1/view

<http://drive.google.com/file/d/1EHDXsLJiF7HeqHJhwq2RIFdPBGZjitzl/view>

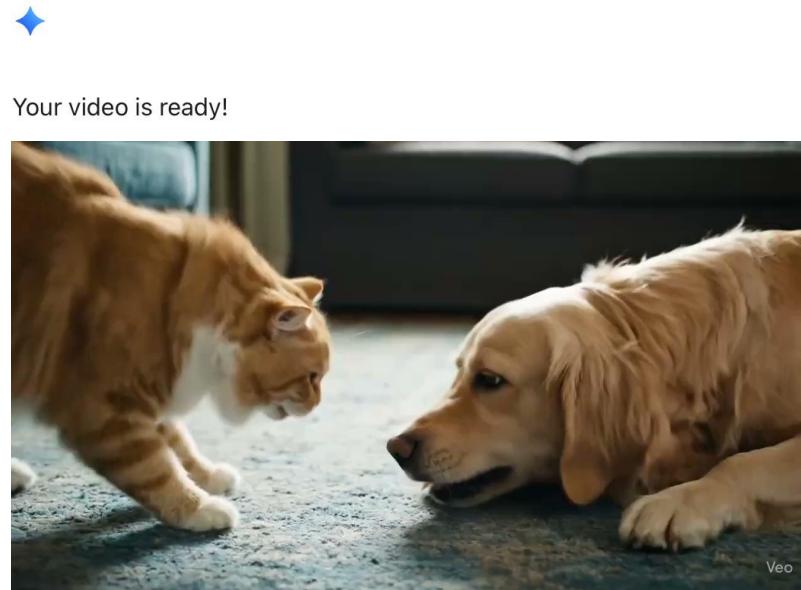




+ 셀프 카메라로 브이로그 촬영한 느낌.
이집트 기자 피라미드 꼭대기에 서서 카메라를 보고 손을 흔드는 남자 한국인 유튜버.
남자 뒤로 피라미드가 보인다

Video 9:16 480p 5s 4v ?

고양이와 강아지가 투닥거리는 영상 3초정도 만들어줘



▶ ▶ ▶

DeepSeek

DeepSeek (深度求索, Hangzhou DeepSeek Artificial Intelligence Basic Technology Research Co., Ltd.) is an AI startup headquartered in Hangzhou, Zhejiang Province, China. It was founded on July 17, 2023, by Liang Wenfeng and is backed by Chinese hedge fund High-Flyer as its main investor and owner.

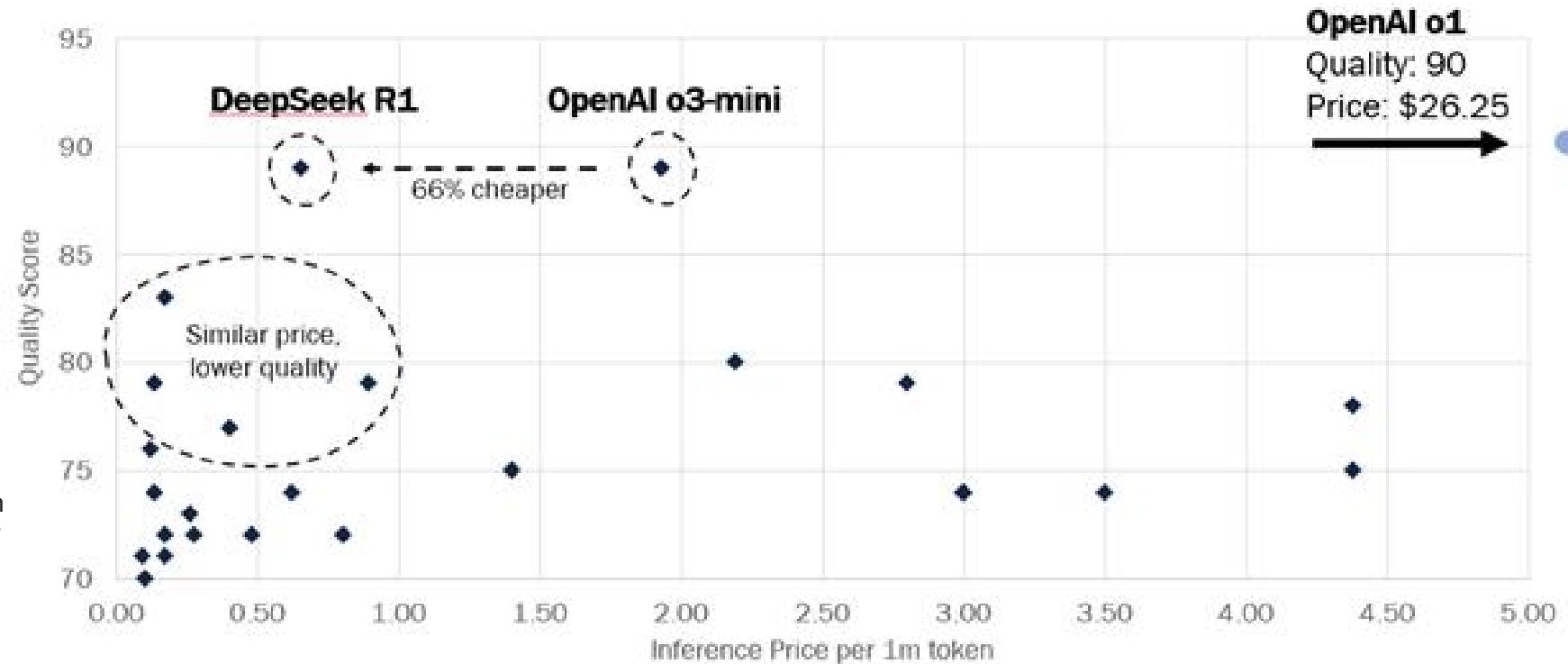


Taking Stock of the DeepSeek Shock February 5, 2025

On January 27, the U.S. stock market and tech stocks [took one of the biggest tumbles](#) in history, with chipmaker Nvidia falling 18%, losing \$589 billion in market value.



Liang Wenfeng (Born 1985)
DeepSeek's chief executive
Graduated from [Zhejiang University](#),
Liang received a [Bachelor of Engineering](#) in electronic information engineering in 2007 and a [Master of Engineering](#) in information and communication engineering in 2010.



DeepSeek is making a notable impact on the market by offering high-quality AI inference at highly competitive rates. The blended cost [1](#) of the [R1 model is \\$0.65 per 1 million tokens](#), which is substantially lower than [the \\$26.25 for OpenAI's o1 model](#) while [maintaining comparable quality](#).



MOONSHOT AI - KIMI K2

Chinese AI unicorn company Moonshot's newly unveiled model, Kimi K2, is attracting attention for its outstanding performance. Some of its features are said to outperform OpenAI's Chat GPT, Google Gemini, and China's Deepseek, leading to speculation that "another Deepseek moment that shook the global AI industry is coming again."

July 2025.

“전 세계 AI 업계를 흔든 딥시크 충격이 또다시 왔다.”

중국의 인공지능(AI) 유니콘 기업 문샷이 새로 공개한 모델 ‘키미 K2’가 뛰어난 성능을 보이며 화제를 모으고 있다. 일부 기능은 미국 오픈AI의 챗 GPT, 구글 제미나이, 중국의 딥시크보다 나은 성능을 보인다는 평가를 받으면서 “전 세계 AI 업계를 흔든 딥시크 모먼트가 또다시 오는 것 아니냐”는 얘기까지 나온다.

18일 국제 학술지 네이처에 따르면, 문샷이 지난 11일 새롭게 공개한 AI 모델 키미 K2는 다양한 AI 모델 능력 평가에서 높은 점수를 받았다. 가장 탁월한 것은 코딩 능력이다. AI의 코딩 능력을 평가하는 ‘라이브코드벤치’에서 키미 K2는 ’53.7% 패스@1’ 점수를 받았다. 단 한 번의 시도로 코딩 과제의 53.7%를 정확히 해결했다는 뜻이다. GPT-4.1(44.7%)나, 딥시크-V3(46.9%)를 뛰어넘는 점수다. 수학 문제 500개를 얼마나 논리적으로 푸는지 보는 ‘매스-500’ 능력 평가에선 97.4%의 정답률을 보여 구글 제미나이 2.5프로(95.2%), 앤트로피의 클로드 3.7 소넷(96.2%)를 제치고 1위에 올랐다.

중국발 AI 충격, 이번엔 '키미 K2'

문샷 새 모델 코딩·수학 능력, 챗GPT·제미나이·딥시크 능가

송혜진 기자

업데이트 2025.07.21. 09:46 ▾



Yang Zhilin in Beijing



중국 AI 기업 문샷이 지난 11일 새 AI모델 '키미K2' 공개했다. /MOONSHOT AI

The company's founder and CEO, 31-year-old Yang Zhilin (楊植麟), graduated from Tsinghua University's computer science department, earned his Ph. D. from Carnegie Mellon, and worked for Meta's AI team and Google Brain in the United States.



AI 기반 인테리어 디자인 플랫폼으로, 사용자가 공간 사진이나 도면을 업로드하면 **포토리얼한 3D 시각화**를 빠르게 생성해줍니다.

실내뿐 아니라 외부 공간(외관, 조경 등) 디자인도 지원하며, 정밀한 랜더링과 스타일로 공간을 구현할 수 있습니다.

Sketch-to-Render: 단순 스케치나 도면으로부터 3D 랜더 이미지 생성
가구 배치, 조명, 색상 변경 등 상호작용 편집 기능 제공

Virtual Staging: 빈 공간에 가구를 채우고 스타일에 따라 연출 가능

ReRoom AI는 인공지능을 활용해 공간 사진, 스케치, 또는 3D 모델을 업로드하면 몇 초 만에 포토리얼한 인테리어 디자인 시각화를 제공하는 웹 기반 플랫폼입니다.

사진·스케치·3D 모델 입력 지원: 사용자는 자신의 공간을 JPG, PNG, 렌더 이미지 등 다양한 형태로 업로드할 수 있습니다.

20가지 이상의 디자인 스타일 제안: 미니멀, 모던, 보헤미안, 스칸디나비아 테마 등 다양한 스타일 중 선택 가능하며, 스타일 조합이 가능합니다.

맞춤형 조절 옵션과 모드: 조명, 색상 구성, 질감, 시간대 설정 등을 조절할 수 있으며, Precise / Balance / Creative 모드로 표현 강도를 선택할 수 있습니다.



mnml.ai(발음: “미니멀”)는 건축가와 인테리어 디자이너를 위해 고안된 AI 렌더링 플랫폼입니다. 사진, 스케치, 2D 도면 또는 3D 모델을 업로드하거나 텍스트 프롬프트를 입력하면, 짧은 시간 안에 고퀄리티의 시각화를 생성할 수 있는 다양한 기능을 제공합니다. 주요 특징은 다음과 같습니다:

Exterior AI, Interior AI, Landscape AI, Virtual Staging, Sketch-to-Image, Masterplan AI, Style Transfer, Concept Statement Generator, Edit & Modify (Canvas), Render Enhancer, Video AI 등 12개 이상의 도구와 40개 이상의 스타일을 지원합니다

이집트 카이로(Cairo, Egypt) 2020년 설립 [CB InsightsTracxn](#)

Luw.ai(루우에이아이)는 AI 기반 인테리어 및 외부 공간 디자인 플랫폼으로, 누구나 쉽게 공간을 시각화하고 리디자인할 수 있도록 돕습니다. 아래는 주요 특징을 정리한 내용입니다.

Luw.ai는 사진, 스케치 또는 3D 모델(스크린샷 포함)을 업로드하면, 몇 초 만에 포토리얼 한 인테리어 및 외부 공간 렌더를 생성해주는 AI 도구입니다

건축가, 인테리어 디자이너, 부동산 전문가, DIY 사용자 등 전문가부터 초보자까지 폭넓게 활용할 수 있도록 설계되었습니다

사용자 취향과 스타일에 맞춰 학습하는 맞춤형 AI 페르소나 생성 기능



The screenshot shows the Canva interface. On the left, there's a sidebar with categories like 'Graphic Design', 'Video & Animation', 'Marketing', 'Photography', 'Design', 'Business', and 'Personal'. The main area features a 'GIF로 표현하세요' (Express with GIF) section with a search bar and a grid of GIFs. Below it is a '추천 앱' (Recommended Apps) section with cards for 'Typeform', 'Choppy Crop', '3D Maker', 'Pixelify', and 'Flodesk'. At the bottom, there's a '인기 앱' (Popular App) section with a card for 'Canva Pro'.

The screenshot shows the InstantInterior AI website. It features a header with 'LAUNCH SPECIAL: Only 5 spots left at 50% off!' and navigation links for 'How It Works', 'Pricing', and 'Blog'. The main content area has a heading 'Transform Any Space with AI-Generated Interior Design' and a subtext 'Create stunning AI interior designs in minutes. Just upload photos and get professional results.' Below this is a 'How Our AI Interior Design Generator Works' section with three steps: 'Step 1: Upload room photos', 'Step 2: Describe your desired style', and 'Step 3: Get your designs'. There's also a 'Upload Room Image' button and dropdown menus for 'Room Type' (Living Room) and 'Style' (Modern). To the right, there's an 'upload a picture of the space' button with an arrow pointing to a photo of a kitchen, followed by a '100% AI generated' image of a modern living room.

캔바(Canva)는 사용자들이 소셜 미디어 그래픽, 프레젠테이션, 포스터, 문서 등 다양한 시각 콘텐츠를 쉽게 만들 수 있도록 돋는 호주의 온라인 그래픽 디자인 플랫폼입니다. 템플릿을 기반으로 하며, 개인 무료 이용과 Canva Pro 및 Enterprise 등 유료 구독 서비스를 제공합니다. 사람들에게 힘이 되는 기업을 추구하며, 정직하고 개방적이며 전설적인 태도를 중요하게 생각합니다. 2012년 설립되었으며, 2021년에는 월 활성 사용자 6천만 명을 기록하며 빠르게 성장한 디자인 SaaS(Software as a Service) 툴 기업입니다. 프레젠테이션, 문서, 포스터, 동영상 등 다양한 템플릿을 제공하여 누구나 쉽게 디자인할 수 있도록 돋습니다.

InstantInterior AI는 인공지능 기반의 실시간 실내 디자인 생성 플랫폼입니다. 사용자는 사진을 업로드하고 원하는 스타일을 설명하면, AI가 몇 분 만에 포토리얼한 인테리어 디자인과 함께 가구 및 소품 추천 리스트를 제공합니다.

"4 Best Free AI Interior Design Apps in 2025" 리뷰에서도 InstantInterior AI는 단연 에디터 추천 1위로 꼽힙니다. 무료 이용 시에도 고해상도 디자인 5개 제공, 워터마크 없음, 스타일 및 가구 추천 등을 동일하게 받을 수 있다는 점이 매우 돋보입니다.

본 보고서는 실거주 중인 본인의 방 이미지를 인풋 데이터로 제공하고, 동일한 공통 프롬프트("인테리어를 일본풍으로 랜더링해 줘")를 각 인테리어 생성형 AI 사이트에 입력하여, 각 사이트에서 제공하는 랜더링 이미지를 비교·분석한 결과와 소감을 정리한 것입니다.

공통 입력 프롬프트
인테리어를 일본풍으로 랜더링해 줘

여러 사이트를 비교해본 결과, instantinteriorai.com과 reroom.ai의 사용성과 랜더링 결과가 가장 만족스러웠습니다. 두 사이트 모두 실제 인테리어에 참고할 수 있을 정도로 품질이 높았고, 사용법도 직관적이었습니다. 반면, paintit.ai, mnml.ai, luw.ai는 프롬프트 작성의 어려움, 현실감 부족, 불필요한 요소 등으로 인해 만족도가 낮았습니다.



reroom.ai

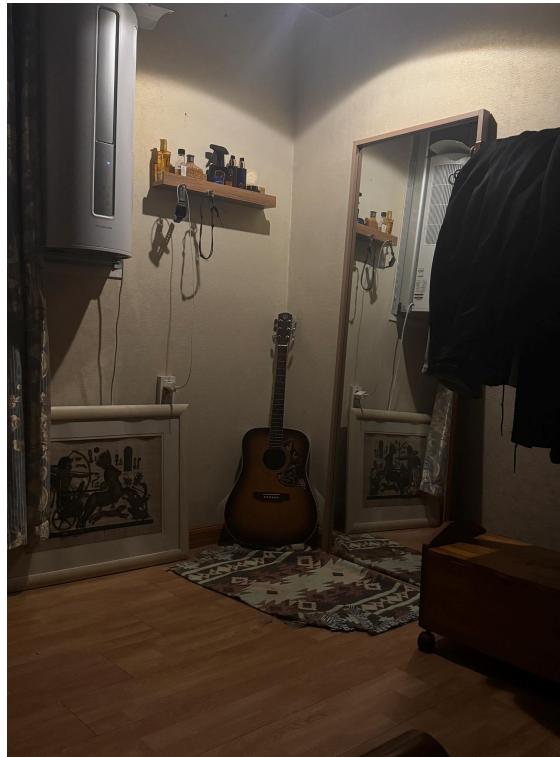


paintit.ai

본 보고서는 실거주 중인 본인의 방 이미지를 인풋 데이터로 제공하고, 동일한 공통 프롬프트("인테리어를 일본풍으로 랜더링해 줘")를 각 인테리어 생성형 AI 사이트에 입력하여, 각 사이트에서 제공하는 랜더링 이미지를 비교·분석한 결과와 소감을 정리한 것입니다.

공통 입력 프롬프트
인테리어를 일본풍으로 랜더링해 줘

여러 사이트를 비교해본 결과, instantinteriorai.com과 reroom.ai의 사용성과 랜더링 결과가 가장 만족스러웠습니다. 두 사이트 모두 실제 인테리어에 참고할 수 있을 정도로 품질이 높았고, 사용법도 직관적이었습니다. 반면, paintit.ai, mnml.ai, luw.ai는 프롬프트 작성의 어려움, 현실감 부족, 불필요한 요소 등으로 인해 만족도가 낮았습니다.



mnml.ai

luw.ai

instantinteriorai.com

NVIDIA GeForce RTX 4080

\$1,600~ 2,500,000원~

AD103

GRAPHICS PROCESSOR

9728

CORES

304

TMUS

112

ROPs

16 GB

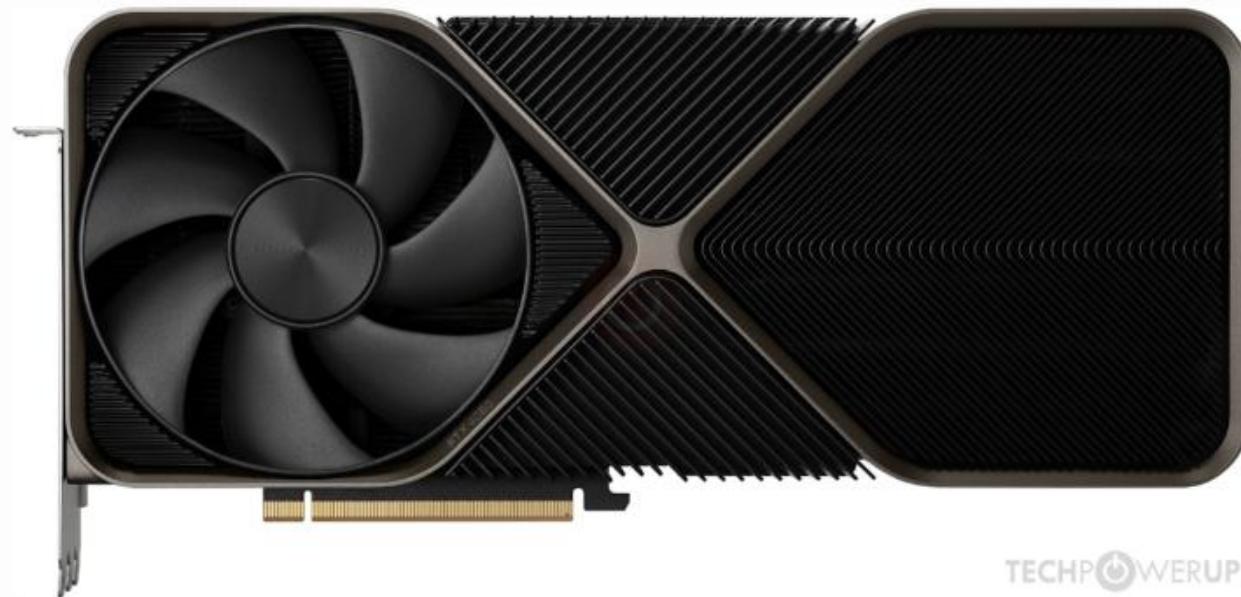
MEMORY SIZE

GDDR6X

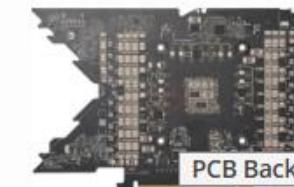
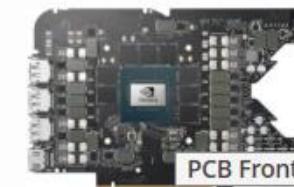
MEMORY TYPE

256 bit

BUS WIDTH



TECHPOWERUP





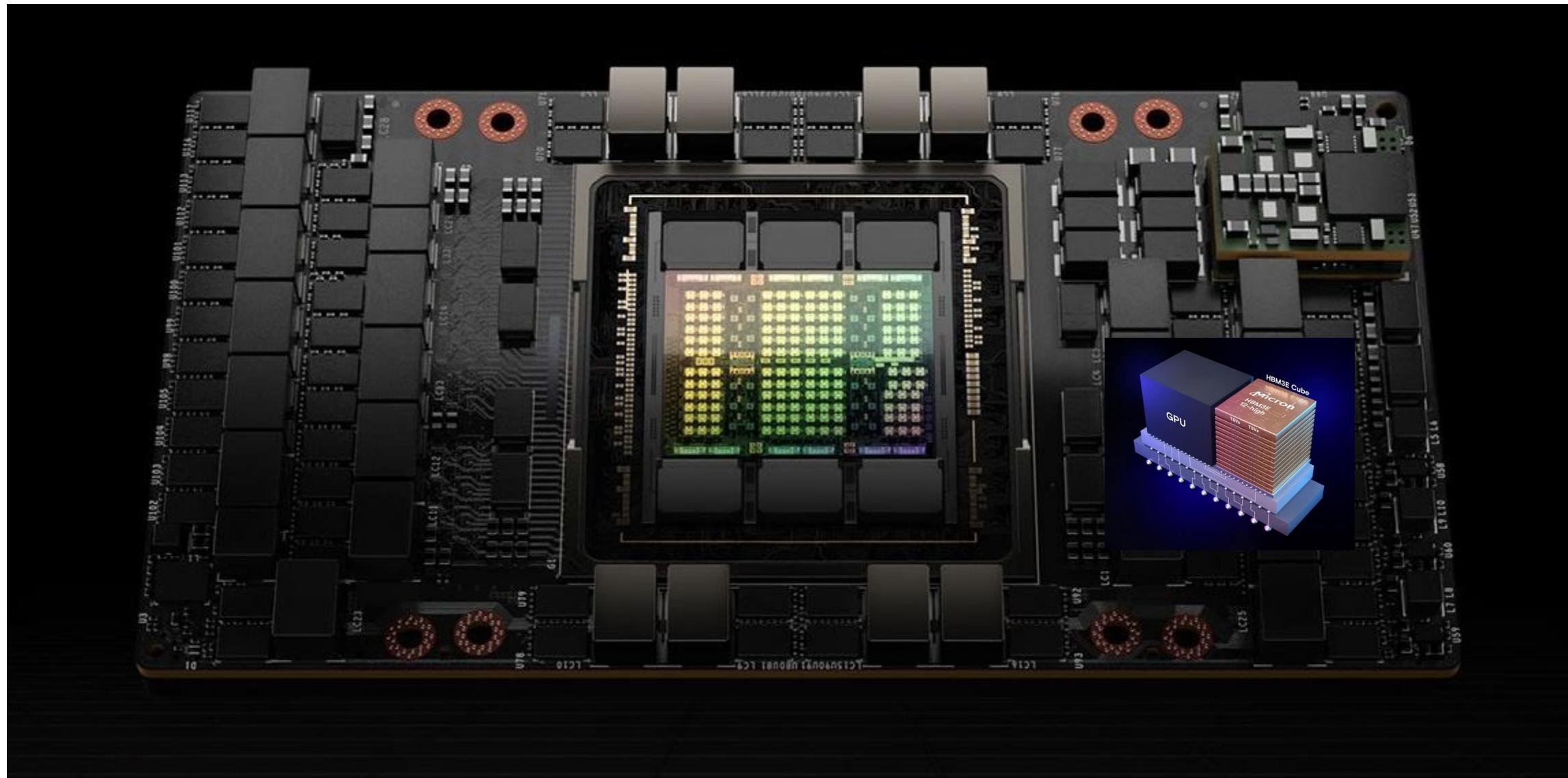
| Lineup | GeForce | Tesla | Quadro |
|------------|---------------|---------------|---------------|
| Name | RTX 2080 Ti | V100 | RTX 8000 |
| Generation | 7.5 (Turing) | 7.0 (Volta) | 7.5 (Turing) |
| Picture | | | |
| FP32 | 11,750 GFLOPS | 14,899 GFLOPS | 16,300 GFLOPS |
| FP64 | 367 GFLOPS | 7,450 GFLOPS | 510 GFLOPS |
| Memory | 11GB (GDDR6) | 32GB (HBM2) | 48GB (GDDR6) |
| Bandwidth | 616GB/s | 900GB/s | 672GB/s |
| Power | 250W | 250W | 295W |
| MSRP | \$ 999 | \$ 10,000 | \$ 5,500 |

DGX 스파크, 블랙 웨일 울트라가 탑재된 DGX 스테이션
엔비디아 'GB300' 블랙 웨일 울트라 슈퍼칩을 탑재한 첫번째
데스크톱 시스템 소형 AI 서버 784GB의 대용량 메모리
'커넥트엑스-8 슈퍼닉(ConnectX-8 SuperNIC)' 네트워킹 기술
10만달러(약 1억4600만원) 정도로 추정

| VideoCardz.com | NVIDIA H100 | NVIDIA A100 | NVIDIA Tesla V100 | NVIDIA Tesla P100 |
|------------------|---------------------|---------------------|---------------------|---------------------|
| Picture | | | | |
| GPU | GH100 | GA100 | GV100 | GP100 |
| Transistors | 80B | 54.2B | 21.1B | 15.3B |
| Die Size | 814 mm ² | 828 mm ² | 815 mm ² | 610 mm ² |
| Architecture | Hopper | Ampere | Volta | Pascal |
| Fabrication Node | TSMC N4 | TSMC N7 | 12nm FFN | 16nm FinFET+ |
| GPU Clusters | 132/114* | 108 | 80 | 56 |
| CUDA Cores | 16896/14592* | 6912 | 5120 | 3584 |
| L2 Cache | 50MB | 40MB | 6MB | 4MB |
| Tensor Cores | 528/456* | 432 | 320 | - |
| Memory Bus | 5120-bit | 5120-bit | 4096-bit | 4096-bit |
| Memory Size | 80 GB HBM3/HBM2e* | 40/80GB HBM2e | 16/32 HBM2 | 16GB HBM2 |
| TDP | 700W/350W* | 250W/300W/400W | 250W/300W/450W | 250W/300W |
| Interface | SXM5/*PCIe Gen5 | SXM4/PCIe Gen4 | SXM2/PCIe Gen3 | SXM/PCIe Gen3 |
| Launch Year | 2022 | 2020 | 2017 | 2016 |
| | \$35,000 | \$20,000 | | |

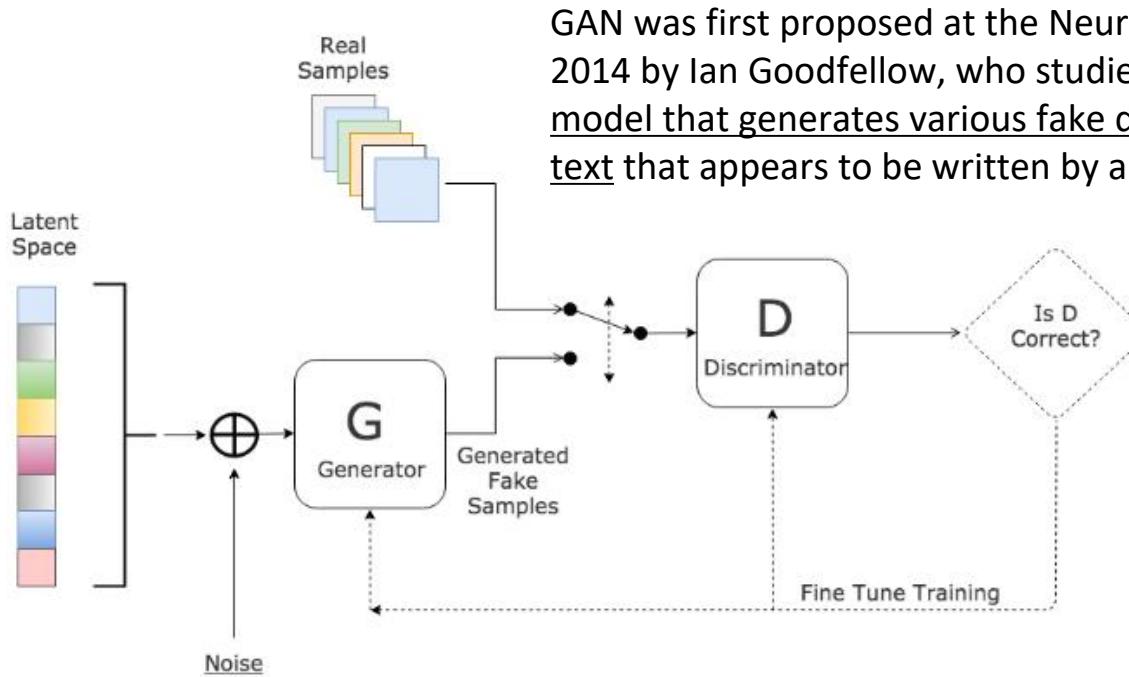
<https://www.nvidia.com/en-us/data-center/h100/> <https://prod.danawa.com/info/?pcode=49261274> 4,000만원

<https://www.aitimes.com/news/articleView.html?idxno=166628>

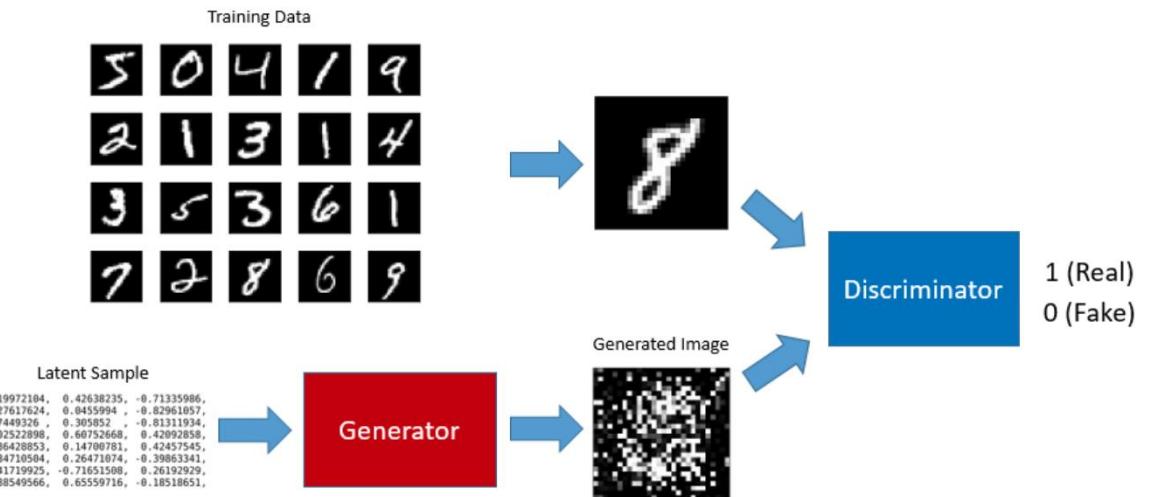


Generative Adversarial Networks

A **generative adversarial network (GAN)** is a class of [machine learning](#) framework and a prominent framework for approaching [generative AI](#).^{[1][2]} The concept was initially developed by [Ian Goodfellow](#) and his colleagues in June 2014.^[3] In a GAN, two [neural networks](#) contest with each other in the form of a [zero-sum game](#), where one agent's gain is another agent's loss. Given a training set, this technique learns to generate new data with the same statistics as the training set. For example, a GAN trained on photographs can generate new photographs that look at least superficially authentic to human observers, having many realistic characteristics. Though originally proposed as a form of [generative model](#) for [unsupervised learning](#), GANs have also proved useful for [semi-supervised learning](#),^[4] fully [supervised learning](#),^[5] and [reinforcement learning](#).^[6]



GAN was first proposed at the Neural Information Processing Systems (NIPS) in 2014 by Ian Goodfellow, who studied machine learning at Google Brain. [GAN](#) is a model that generates various fake data, such as images that are close to reality or text that appears to be written by a person.



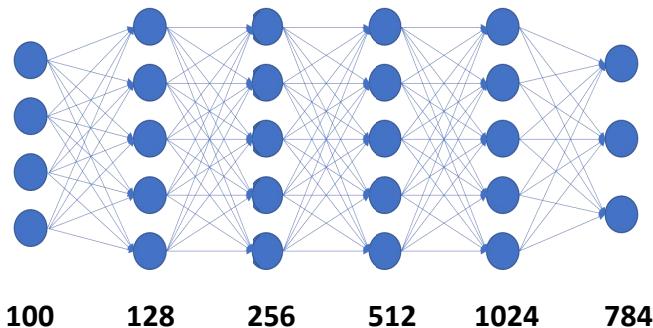
Generative Adversarial Networks

```
latent_dim = 100
```

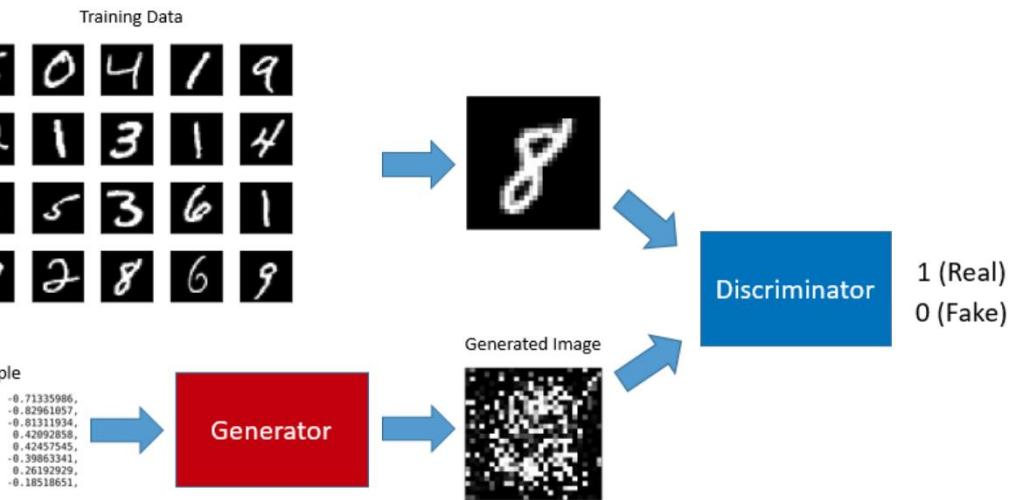
```
# 생성자(Generator) 클래스 정의
class Generator(nn.Module):
    def __init__(self):
        super(Generator, self).__init__()
        # 하나의 블록(block) 정의
        def block(input_dim, output_dim, normalize=True):
            layers = [nn.Linear(input_dim, output_dim)]
            if normalize:
                # 배치 정규화(batch normalization) 수행(차원 동일)
                layers.append(nn.BatchNorm1d(output_dim, 0.8))
            layers.append(nn.LeakyReLU(0.2, inplace=True))
            return layers
    # 생성자 모델은 연속적인 여러 개의 블록을 가짐
    self.model = nn.Sequential(
        *block(latent_dim, 128, normalize=False),
        *block(128, 256),
        *block(256, 512),
        *block(512, 1024),
        nn.Linear(1024, 1 * 28 * 28),
        nn.Tanh()
    )

    def forward(self, z):
        img = self.model(z)
        img = img.view(img.size(0), 1, 28, 28)
        return img
```

$$\begin{array}{r} 1 \ 0 \ 1 \quad x \quad 1 \ 2 \ 8 = \quad 1 \ 2 , \ 9 \ 2 \ 8 \\ 1 \ 2 \ 9 \quad x \quad 2 \ 5 \ 6 = \quad 3 \ 3 , \ 0 \ 2 \ 4 \\ 2 \ 5 \ 7 \quad x \quad 5 \ 1 \ 2 = \quad 1 \ 3 \ 1 , \ 5 \ 8 \ 4 \\ 5 \ 1 \ 3 \quad x \quad 1 \ 0 \ 2 \ 4 = \quad 5 \ 2 \ 5 , \ 3 \ 1 \ 2 \\ 1 \ 0 \ 2 \ 5 \quad x \quad 7 \ 8 \ 4 = \quad 8 \ 0 \ 3 , \ 6 \ 0 \ 0 \end{array}$$



100 128 256 512 1024 784



Generative Adversarial Networks

$$\begin{array}{r} 7 \ 8 \ 5 \ x \ 5 \ 1 \ 2 = 4 \ 0 \ 1 , \ 9 \ 2 \ 0 \\ 5 \ 1 \ 3 \ x \ 2 \ 5 \ 6 = 1 \ 3 \ 1 , \ 3 \ 2 \ 8 \\ 2 \ 5 \ 7 \ x \ 1 \quad \quad \quad = \quad \quad \quad 2 \ 5 \ 7 \end{array}$$

판별자(Discriminator) 클래스 정의

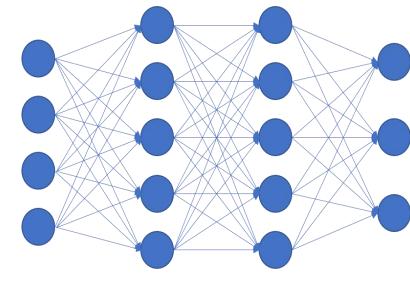
```
class Discriminator(nn.Module):
    def __init__(self):
        super(Discriminator, self).__init__()

        self.model = nn.Sequential(
            nn.Linear(1 * 28 * 28, 512),
            nn.LeakyReLU(0.2, inplace=True),
            nn.Linear(512, 256),
            nn.LeakyReLU(0.2, inplace=True),
            nn.Linear(256, 1),
            nn.Sigmoid(),
        )
```

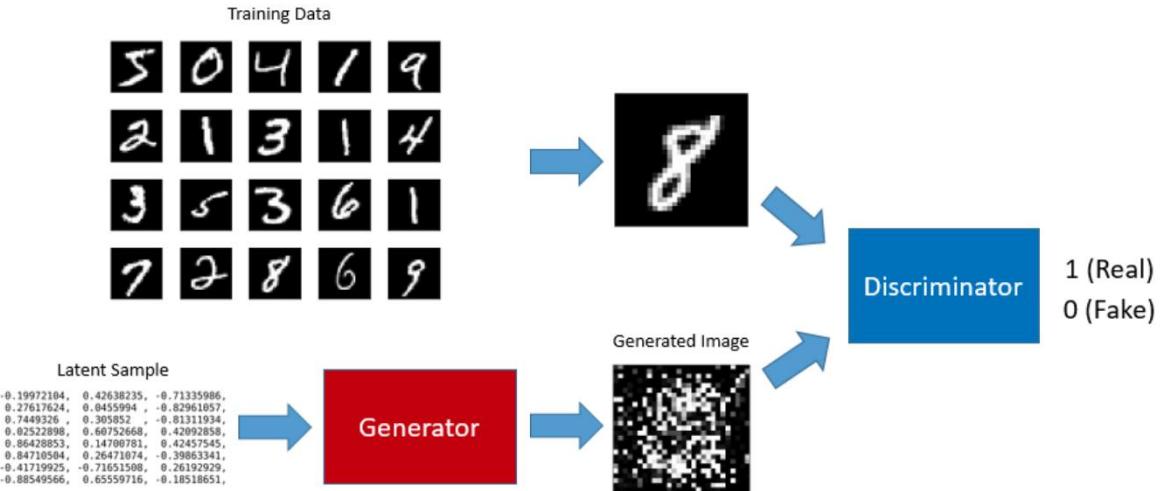
이미지에 대한 판별 결과를 반환

```
def forward(self, img):
    flattened = img.view(img.size(0), -1)
    output = self.model(flattened)

    return output
```



7 8 4 5 1 2 2 5 6 1

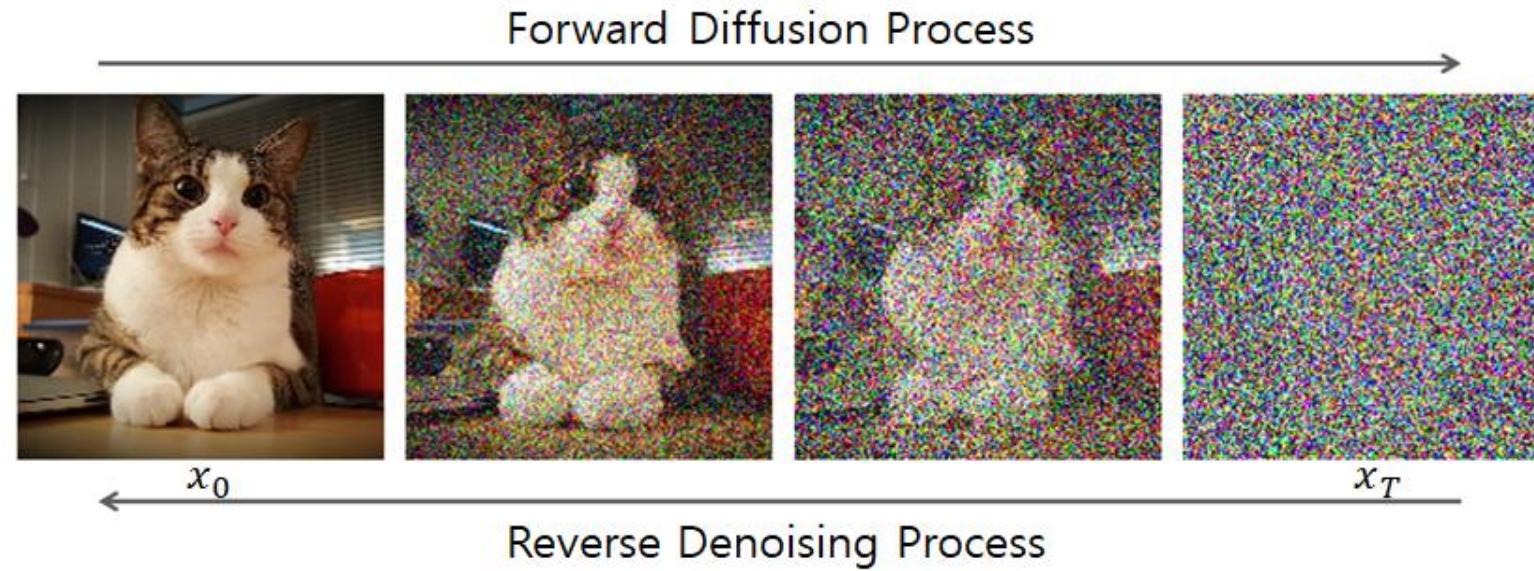


Diffusion Model - MLE

GAN, VAE와 같은 생성 모델(Generative Model) 중 하나로써, 2022년에 이슈가 되었던 text-to-image 모델인 Stable-Diffusion, DALL-E-2, Imagen의 기반이 되는 모델이다.

Diffusion Model 이란?

입력 이미지에 (정규 분포를 가진) Noise를 여러 단계에 걸쳐 추가하고, 여러 단계에 걸쳐 (정규 분포를 가진) Noise를 제거함으로써, 입력 이미지와 유사한 확률 분포를 가진 결과 이미지를 생성하는 모델이다.



Deep Unsupervised Learning using Nonequilibrium Thermodynamics

Jascha Sohl-Dickstein

Stanford University

Eric A. Weiss

University of California, Berkeley

Niru Maheswaranathan

Stanford University

Surya Ganguli

Stanford University

Abstract

A central problem in machine learning involves modeling complex data-sets using highly flexible families of probability distributions in which learning, sampling, inference, and evaluation are still analytically or computationally tractable. Here, we develop an approach that simultaneously achieves both flexibility and tractability. The essential idea, inspired by non-equilibrium

these models are unable to aptly describe structure in rich datasets. On the other hand, models that are *flexible* can be molded to fit structure in arbitrary data. For example, we can define models in terms of any (non-negative) function $\phi(\mathbf{x})$ yielding the flexible distribution $p(\mathbf{x}) = \frac{\phi(\mathbf{x})}{Z}$, where Z is a normalization constant. However, computing this normalization constant is generally intractable. Evaluating, training, or drawing samples from such flexible models typically requires a very expensive Monte Carlo process.

JASCHA@STANFORD.EDU

EAWEISS@BERKELEY.EDU

NIRUM@STANFORD.EDU

SGANGULI@STANFORD.EDU

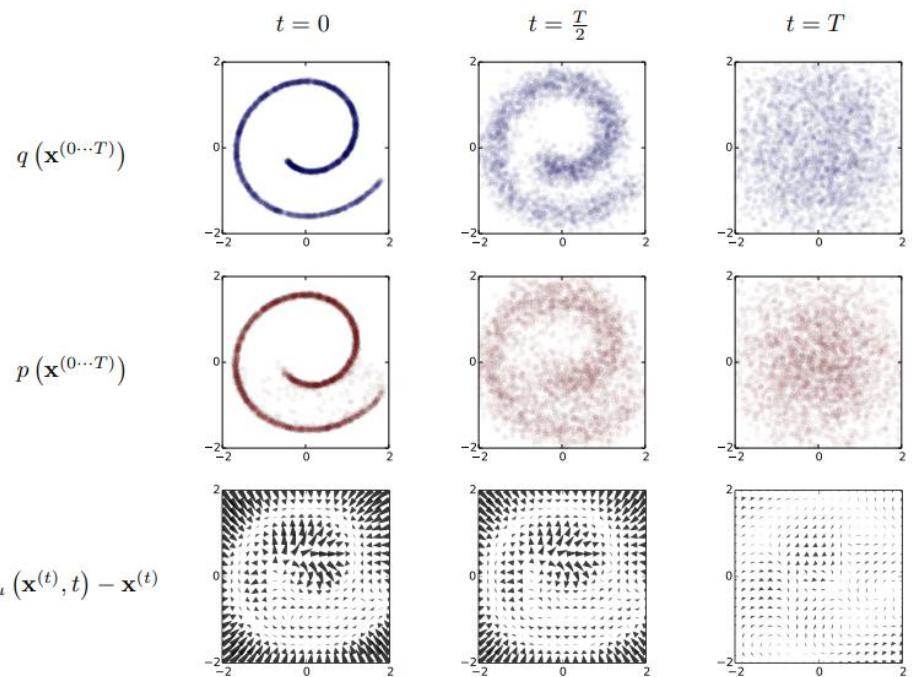


Figure 1. The proposed modeling framework trained on 2-d swiss roll data. The top row shows time slices from the forward trajectory $q(\mathbf{x}^{(0\cdots T)})$. The data distribution (left) undergoes Gaussian diffusion, which gradually transforms it into an identity-covariance Gaussian (right). The middle row shows the corresponding time slices from the trained reverse trajectory $p(\mathbf{x}^{(0\cdots T)})$. An identity-covariance Gaussian (right) undergoes a Gaussian diffusion process with learned mean and covariance functions, and is gradually transformed back into the data distribution (left). The bottom row shows the drift term, $f_\mu(\mathbf{x}^{(t)}, t) - \mathbf{x}^{(t)}$, for the same reverse diffusion process.

Our method uses a Markov chain to gradually convert one distribution into another, an idea used in non-equilibrium statistical physics (Jarzynski, 1997) and sequential Monte Carlo (Neal, 2001). We build a generative Markov chain which converts a simple known distribution (e.g. a Gaussian) into a target (data) distribution using a diffusion process. Rather than use this Markov chain to approximately evaluate a model which has been otherwise defined, we explicitly define the probabilistic model as the endpoint of the Markov chain. Since each step in the diffusion chain has an analytically evaluable probability, the full chain can also be analytically evaluated. Learning in this framework involves estimating small perturbations to a diffusion process. Estimating small perturbations is more tractable than explicitly describing the full distribution with a single, non-analytically-normalizable, potential function. Furthermore, since a diffusion process exists for any smooth target distribution, this method can capture data distributions of arbitrary form. We demonstrate the utility of these diffusion probabilistic models by training high log likelihood models for a two dimensional swiss roll, binary sequence, handwritten digit (MNIST), and several natural image (CIFAR-10, bark, and dead leaves) datasets.

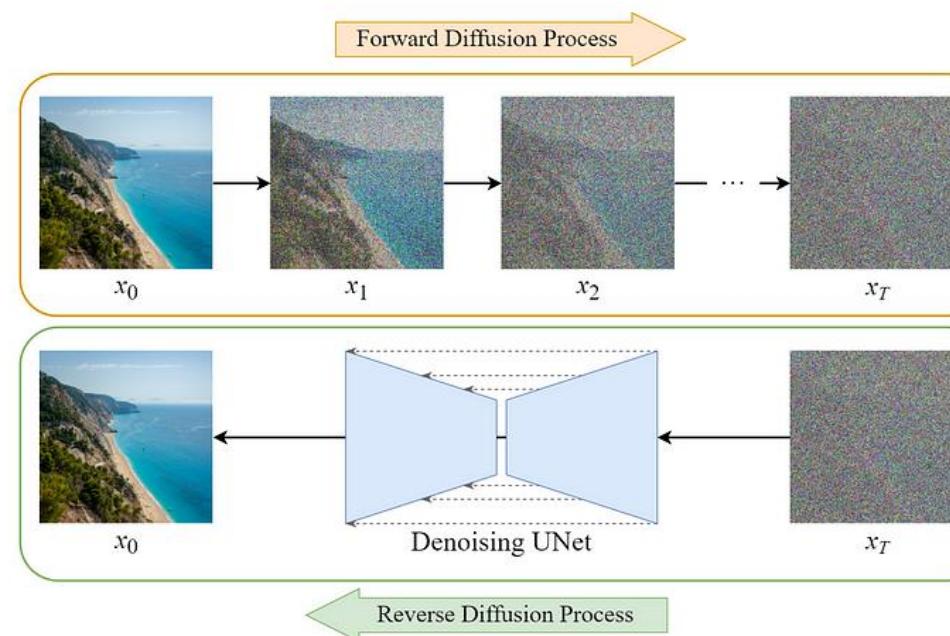
Diffusion Model

입력 이미지에서 Noise가 서서히 확산하기 때문에 Diffusion이라는 이름이 붙었다고 합니다.

Forward Diffusion Process에서는 이미지에 고정된(fixed) 정규 분포(=Gaussian분포)로 생성된 Noise가 더해지고,

Reverse Diffusion Process에서는 이미지를 학습된(learned) 정규 분포로 생성된 Noise이미지로 뺍니다.

Diffusion Model이 풀려고 하는 문제는, Forward -> Reverse 단계를 거친 '결과 이미지'를 '입력 이미지'의 확률 분포와 유사하게 만드는 것입니다. 이를 위해 Reverse단계에서, Noise 생성 확률 분포 Parameter인 평균과 표준편차를 업데이트하며 학습이 진행됩니다.. 최초 Diffusion Model(2015년)은 Maximum Log-likelihood Estimation(MLE)을 사용했고, 후속 논문 DDPM(2020년, Denosing Diffusion Probabilistic Model)은 CNN계열의 UNet모델을 사용했다.



Diffusion Model - UNet

Denoising Diffusion Probabilistic Models

Jonathan Ho

UC Berkeley

jonathanho@berkeley.edu

Ajay Jain

UC Berkeley

ajayj@berkeley.edu

Pieter Abbeel

UC Berkeley

pabbeel@cs.berkeley.edu

Abstract

We present high quality image synthesis results using diffusion probabilistic models, a class of latent variable models inspired by considerations from nonequilibrium thermodynamics. Our best results are obtained by training on a weighted variational bound designed according to a novel connection between diffusion probabilistic models and denoising score matching with Langevin dynamics, and our models naturally admit a progressive lossy decompression scheme that can be interpreted as a generalization of autoregressive decoding. On the unconditional CIFAR10 dataset, we obtain an Inception score of 9.46 and a state-of-the-art FID score of 3.17. On 256x256 LSUN, we obtain sample quality similar to ProgressiveGAN. Our implementation is available at <https://github.com/hojonathanho/diffusion>.

Deep generative models of all kinds have recently exhibited high quality samples in a wide variety of data modalities. Generative adversarial networks (GANs), autoregressive models, flows, and variational autoencoders (VAEs) have synthesized striking image and audio samples [14, 27, 3, 58, 38, 25, 10, 32, 44, 57, 26, 33, 45], and there have been remarkable advances in energy-based modeling and score matching that have produced images comparable to those of GANs [11, 55].

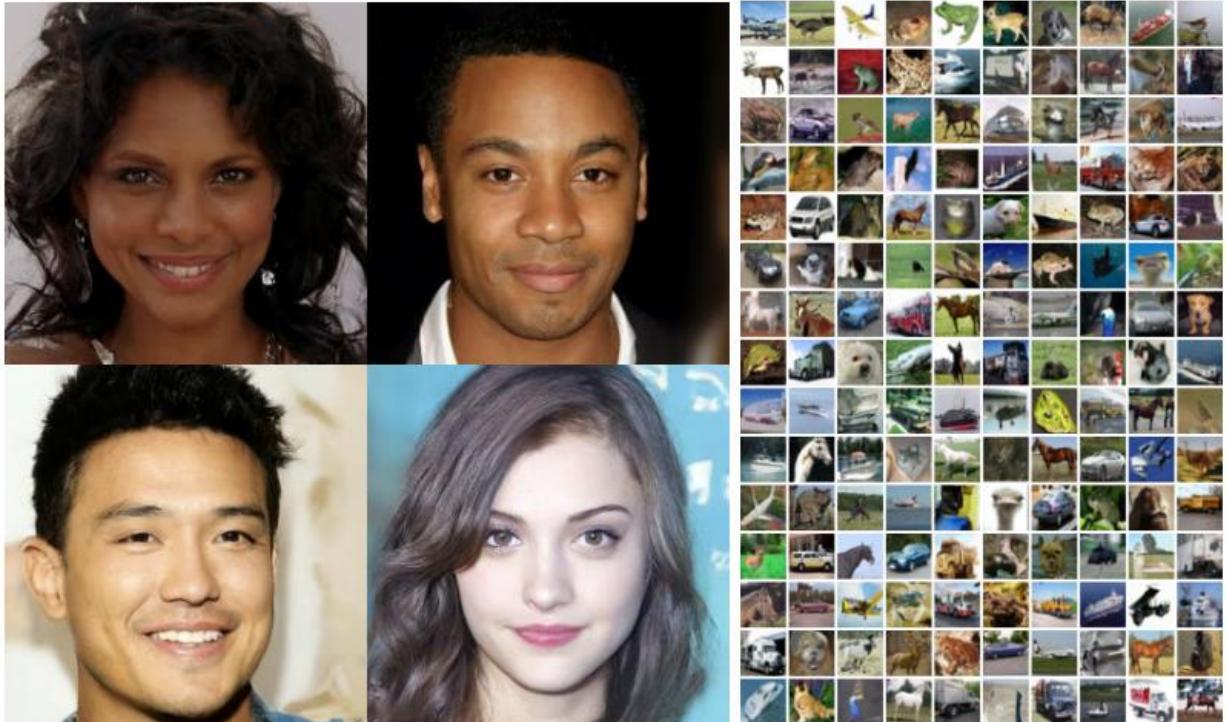
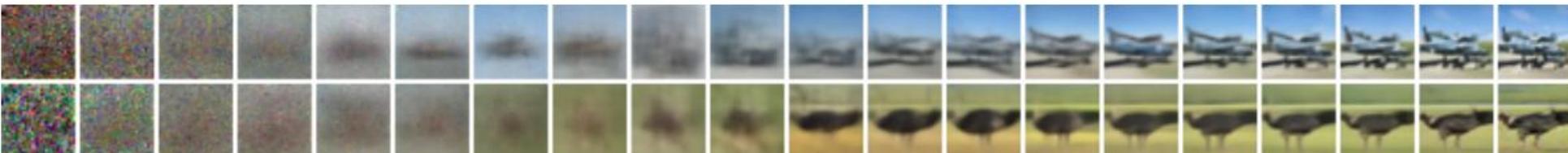


Figure 1: Generated samples on CelebA-HQ 256 × 256 (left) and unconditional CIFAR10 (right)

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.



Stable Diffusion

1. 개요

- Stable Diffusion은 2022년 Stability AI가 공개한 **Latent Diffusion Model (LDM)** 기반의 오픈소스 이미지 생성 모델
- 텍스트 프롬프트를 입력하면 해당 설명에 맞는 이미지를 생성할 수 있음.
- 누구나 PC에서 실행할 수 있을 정도로 경량화되어 있어 Midjourney, DALL·E 같은 클라우드 기반 모델과 달리 자유롭게 커스터마이즈 가능.

2. 핵심 아이디어 – Latent Diffusion

기존 Diffusion Model은 픽셀 공간에서 직접 노이즈를 제거했기 때문에 계산량이 매우 컸다. SD는 다음과 같은 최적화를 적용했음

- VAE(Variational Autoencoder)를 사용하여 이미지를 잠재 공간(latent space)으로 인코딩
- 잠재 공간에서 Diffusion을 수행 → 연산량과 메모리 사용량 크게 감소
- 최종 단계에서 다시 디코딩해 원본 이미지로 변환 즉, Stable Diffusion은 "압축된 이미지 표현" 위에서 Diffusion을 수행하므로 빠르고 효율적

3. 아키텍처 구성

Stable Diffusion은 크게 세 부분으로 나눌 수 있다.

1. Text Encoder

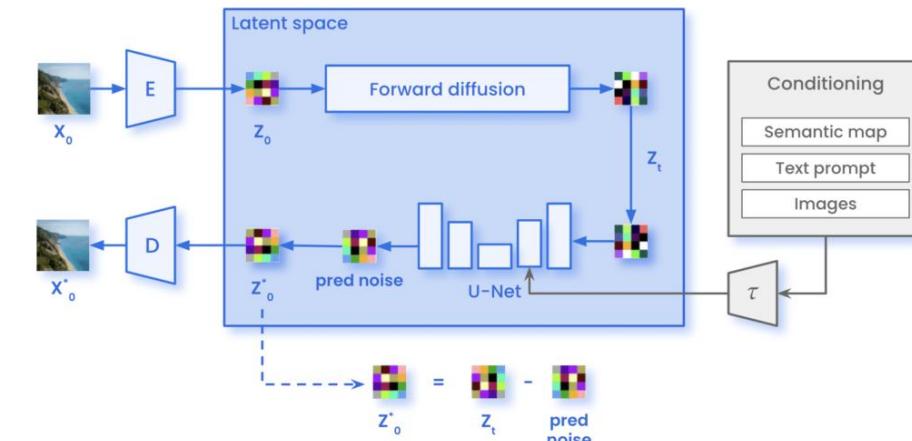
- CLIP의 텍스트 인코더를 사용
- 텍스트 프롬프트를 임베딩 벡터로 변환

2. U-Net (Denoiser)

- 잠재 공간에서 노이즈를 점진적으로 제거하는 핵심 모델
- cross-attention을 통해 텍스트 조건(프롬프트) 반영

3. VAE (Decoder)

- 마지막에 잠재 표현을 다시 픽셀 공간으로 디코딩
- 고해상도 이미지를 출력



High-Resolution Image Synthesis with Latent Diffusion Models

Robin Rombach^{1,*} Andreas Blattmann^{1,*} Dominik Lorenz¹ Patrick Esser³ Björn Ommer¹

¹Ludwig Maximilian University of Munich & IWR, Heidelberg University, Germany

<https://github.com/CompVis/latent-diffusion>

Abstract

By decomposing the image formation process into a sequential application of denoising autoencoders, diffusion models (DMs) achieve state-of-the-art synthesis results on image data and beyond. Additionally, their formulation allows for a guiding mechanism to control the image generation process without retraining. However, since these models typically operate directly in pixel space, optimization of powerful DMs often consumes hundreds of GPU days and inference is expensive due to sequential evaluations. To enable DM training on limited computational resources while retaining their quality and flexibility, we apply them in the latent space of powerful pretrained autoencoders. In co

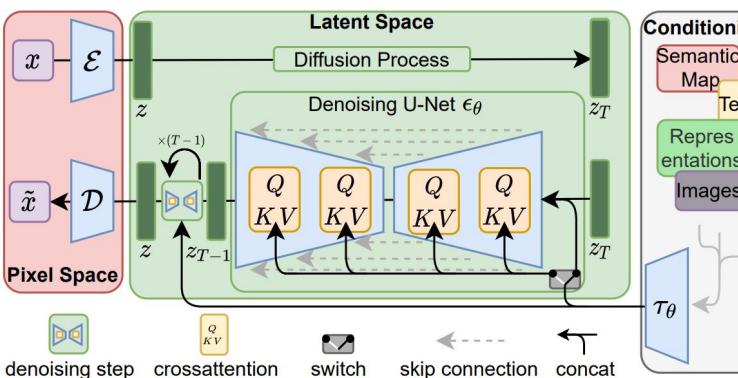


Figure 3. We condition LDMs either via concatenation or by a more general cross-attention mechanism. See Sec. 3.3



Figure 1. Boosting the upper bound on achievable quality with less aggressive downsampling. Since diffusion models offer excellent

need the heavy spatial data in latent space, but the data via suitable representations from the DIV2K [1] and PSNR are 29] and PSNR are 8.

ond [7, 45, 48, 57], conditional image [1]. Moreover, even though they are not suited to tasks such as stroke-based synthesis or generative models, they do not have the difficulties as GANs. In particular, they can

High-Resolution Image Synthesis with Latent Diffusion Models (A.K.A. LDM & Stable Diffusion)

Robin Rombach^{1,2}, Andreas Blattmann^{1,2}, Dominik Lorenz^{1,2}, Patrick Esser³,
Björn Ommer^{1,2}

¹LMU Munich, ²IWR, Heidelberg University, ³Runway
CVPR 2022 (ORAL)



Press Releases



LMU press release



Stable Diffusion in the Press

The original Stable Diffusion model was created in a collaboration with [CompVis](#) and [RunwayML](#) and builds upon the work: [CVPR '22](#)

High-Resolution Image Synthesis with Latent Diffusion Models, Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, Björn Ommer

<https://github.com/Stability-AI/stablediffusion> <https://youtu.be/DaLS4Baiqgk>

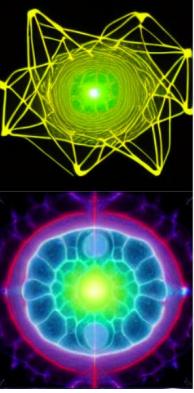
'A zombie in the style of Picasso'

'An image of an animal half mouse half octopus'

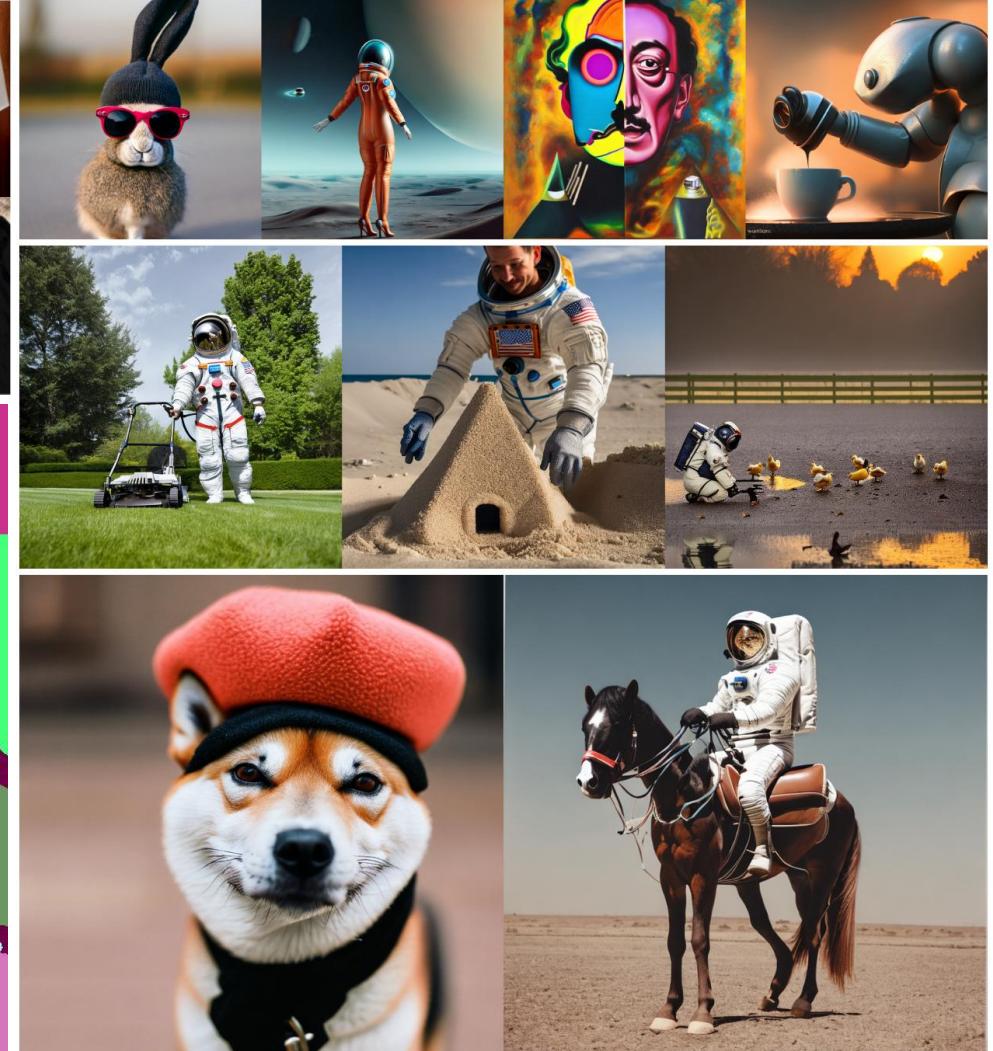
'An illustration of a slightly conscious neural network'

'A painting of a squirrel eating a burger'

'A watercolor painting of a chair that looks like an octopus'
'A shirt with the inscription: "I love generative models!"'



Stable Diffusion Version 2



Stable Diffusion

Stable Diffusion is a [deep learning, text-to-image model](#) released in 2022 based on [diffusion](#) techniques. It is primarily used to generate detailed images conditioned on text descriptions, though it can also be applied to other tasks such as [inpainting](#), [outpainting](#), and generating image-to-image translations guided by a [text prompt](#).^[3] It was developed by researchers from the CompVis Group at [Ludwig Maximilian University of Munich](#) and Runway with a compute donation by Stability AI and training data from non-profit organizations.^{[4][5][6][7]}

Stable Diffusion is a [latent diffusion model](#), a kind of deep generative artificial [neural network](#). Its code and model weights have been released [publicly](#),^[8] and it can run on most consumer hardware equipped with a modest [GPU](#) with at least 8 GB [VRAM](#). This marked a departure from previous proprietary text-to-image models such as [DALL-E](#) and [Midjourney](#) which were accessible only via [cloud services](#).^{[9][10]}

The development of Stable Diffusion was funded and shaped by the start-up company Stability AI.^{[11][10][12][13]} The technical license for the model was released by the CompVis group at Ludwig Maximilian University of Munich.^[10] Development was led by Patrick Esser of Runway and Robin Rombach of CompVis, who were among the researchers who had earlier invented the latent diffusion model architecture used by Stable Diffusion.^[7] Stability AI also credited [EleutherAI](#) and [LAION](#) (a German nonprofit which assembled the dataset on which Stable Diffusion was trained) as supporters of the project.^[7] In October 2022, Stability AI raised US\$101 million in a round led by [Lightspeed Venture Partners](#) and [Coatue Management](#).^[14]

The model was initially trained on the laion2B-en and laion-high-resolution subsets, with the last few rounds of training done on LAION-Aesthetics v2 5+, a subset of 600 million captioned images which the LAION-Aesthetics Predictor V2 predicted that humans would, on average, give a score of at least 5 out of 10 when asked to rate how much they liked them.^{[22][18][23]} The LAION-Aesthetics v2 5+ subset also excluded low-resolution images and images which LAION-5B-WatermarkDetection identified as carrying a [watermark](#) with greater than 80% probability.^[18] Final rounds of training additionally dropped 10% of text conditioning to improve Classifier-Free Diffusion Guidance.^[24] The model was trained using 256 [Nvidia A100](#) GPUs on [Amazon Web Services](#) for a total of 150,000 GPU-hours, at a cost of \$600,000.^{[25][26][27]}

Stable Diffusion



An image generated by Stable Diffusion based on the text prompt "a photograph of an astronaut riding a horse"

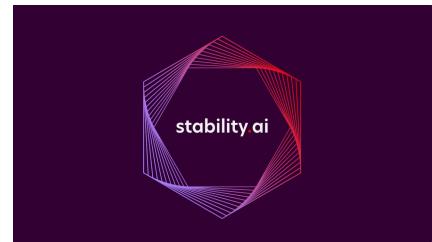
| | |
|---------------------------|--|
| Original author(s) | Runway, CompVis, and Stability AI |
| Developer(s) | Stability AI |
| Initial release | August 22, 2022 |
| Stable release | SDXL 1.0 (model) ^[1] / July 26, 2023 |
| Repository | github.com/Stability-AI/stablediffusion ^[2] |
| Written in | Python ^[2] |
| Operating system | Any that support CUDA kernels |
| Type | Text-to-image model |
| License | Creative ML OpenRAIL-M |
| Website | ommer-lab.com/research/latent-diffusion-models ^[2] |

Stability AI

Stability AI is a developer of an open AI model for images, languages, audio, video, 3D, and biology. These open AI tools are created to help developers design and implement solutions through collective intelligence and augmented technology. Stability AI uses a cluster of more than 4,000 Nvidia A100 GPUs in AWS, which the company uses to train AI systems, including the company's Stable Diffusion. Stability AI CEO Emad Mostaque has suggested it should be capable of training models more efficiently in the future, despite the cost of running this compute environment.

Stability AI's Stable Diffusion is an example of a text-to-image AI. It is an open-source software, differentiating it from other text-to-image AI systems, allowing a community of developers to build on the company's code or power their own commercial offerings. Stability AI offers a commercial version of Stable Diffusion, called DreamStudio, which works to generate revenue by developing the underlying infrastructure and customizing versions of the software for corporate clients.

The DreamStudio website is designed to enable anyone to access creative tools without the need for software installation, coding knowledge, or heavy-duty local GPUs, but comes with a cost. The cost is in DreamStudio credits, in which users are charged one credit per image. Depending on the image resolution and step count users choose, the cost-per-image can go as low as 0.2 credits to as high as 28.2 credits per image. First-time users are offered free credits, and once those credits run out, the user will have to purchase more.



@ [stability.ai](#)



Is a

- Company
- Organization

Company attributes

Industry

- Artificial Intelligence (AI)
- Machine learning
- Application programming interface...
- Developer Tools
- Generative AI
- Technology
- Deep learning

Location

- United Kingdom
- San Francisco (3)
- Notting Hill (3)
- England

LAION

Backend url: <https://splunk>
Index: laion_400m_128G

cute cat

Clip retrieval works by converting the text query to a CLIP embedding, then using that embedding to query a kNN index of clip image embeddings

Display captions Display full captions Display similarities Safe mode Hide duplicate urls Search over [image](#)

This UI may contain results with nudity and is best used by adults. The images are under their own copyright.

Are you seeing near duplicates? KNN search are good at spotting those, especially so in large datasets.

LAION-400M
An open dataset containing 400 million English image-text pairs.

LAION-5B
A dataset consisting of 5.85 billion multilingual CLIP-filtered image-text pairs.

Clip H/14
The largest CLIP (Contrastive Language-Image Pre-training) vision transformer model.

LAION-Aesthetics
A subset of LAION-5B filtered by a model trained to score aesthetically pleasing images.

LAION

PROJECTS

DATASETS

- LAION-400M**
image/text
Status: Released
- LAION5B**
image/text
Status: Released
- Laion-coco**
image/text
Status: Released
- Laion translated**
image/text
Status: Released
- Clip H/14**
image/text
Status: Released

LAION

TEAM

Projects
Team
Blog
Notes
Press
About
FAQ
Donations
Privacy Policy
Dataset Requests
Impressum

Christoph Schuhmann
Organizational Lead / Founder
Master in Physics & Computer Science. Producer of Schools of Trust.

Jenia Jitsev
Scientific Lead / Founder
Senior researcher. Leads Scalable Learning and Multi-Purpose AI (SLAMPAI) Lab at Juelich Supercomputing Center (JSC). PhD in Computer Science, with background in Neuroscience & Machine Learning.

Richard Vencu
Engineering Lead / Founder
AI, automation, and electronics engineer with 28 years of experience in the industry.

Romain Beaumont
Open source contributor
I like to apply scale and deep learning to build AI apps and models.

Robert Kaczmarszyk
Community / Medical Lead / Founder
Technical University of Munich. 3 years experience in epidemiological research.

Theo Coombes
Founding Member
Programmer with a great passion for big data and machine learning.



LAION

Large-scale Artificial Intelligence Open Network

TRULY OPEN AI. 100% NON-PROFIT. 100% FREE.

LAION, as a non-profit organization, provides datasets, tools and models to liberate machine learning research. By doing so, we encourage open public education and a more environment-friendly use of resources by reusing existing datasets and models.



Christoph Schuhmann outside the Gymnasium Klosterschule school in Hamburg, Germany. Photographer: Maria Feck/Bloomberg

Technology | AI

The Future of AI Relies on a High School Teacher's Free Database

With over five billion images, LAION has become central to the future of artificial intelligence — and a growing debate over how to regulate it.

By [Marissa Newman](#) and [Aggi Cantrill](#)

2023년 4월 24일 오후 12:00 GMT+9

LAION



LAION

Large-scale Artificial Intelligence Open Network

TRULY OPEN AI. 100% NON-PROFIT. 100% FREE.

LAION, as a non-profit organization, provides datasets, tools and models to liberate machine learning research. By doing so, we encourage open public education and a more environment-friendly use of resources by reusing existing datasets and models.

LAION (acronym for **L**arge-**s**cale **A**rtificial **I**ntelligence **O**pen **N**etwork) is a German non-profit which makes open-sourced [artificial intelligence](#) models and datasets.^[1] It is best known for releasing a number of large datasets of images and captions scraped from the web which have been used to train a number of high-profile [text-to-image models](#), including [Stable Diffusion](#) and [Imagen](#).^{[2][3]}

In February 2023, LAION was named in the [Getty Images](#) lawsuit against [Stable Diffusion](#) as a non-party.^[4] In April 2023, LAION was directly sued by a German photographer who wanted to have his images removed from the training set.^[5]

On April 15, 2023, LAION and contributors released to public an open source AI assistant chatbot [OpenAssistant](#).

Image datasets

LAION has publicly released a number of large datasets of image-caption pairs which have been widely used by AI researchers. The data is derived from the [Common Crawl](#), a dataset of scraped web pages. The developers searched the crawled html for tags and treated their [alt attributes](#) as captions. They used [CLIP](#) to identify and discard images whose content did not appear to match their captions.^[6] LAION does not host the content of scraped images themselves; rather, the dataset contains [URLs](#) pointing to images, which researchers must download themselves.^[7]

The first such dataset, LAION-400M, was released in August 2021 and consisted of 400 million image-caption pairs. The pairs were extracted from a random subset of webpages scraped by Common Crawl between 2014 and 2021.^[8] It was an attempt to recreate the process used by OpenAI to collect the 400 million image-caption pairs they used to train the CLIP model - the company had chosen to open-source the model's code and weights, but not its training dataset.^[6] [Imagen](#), a text-to-image model announced by Google Brain in 2022, was trained on LAION-400M in combination with private internal datasets.^[9]

WebUI

2022.08

Stable Diffusion WebUI는 **Stable Diffusion** 이미지 생성 모델을 쉽고 편하게 사용할 수 있도록 만들어진 웹 기반 인터페이스입니다. 가장 널리 사용되는 것은 **AUTOMATIC1111 WebUI**로, 전 세계에서 커뮤니티 중심으로 활발히 개발되고 있는 오픈소스 프로젝트입니다.

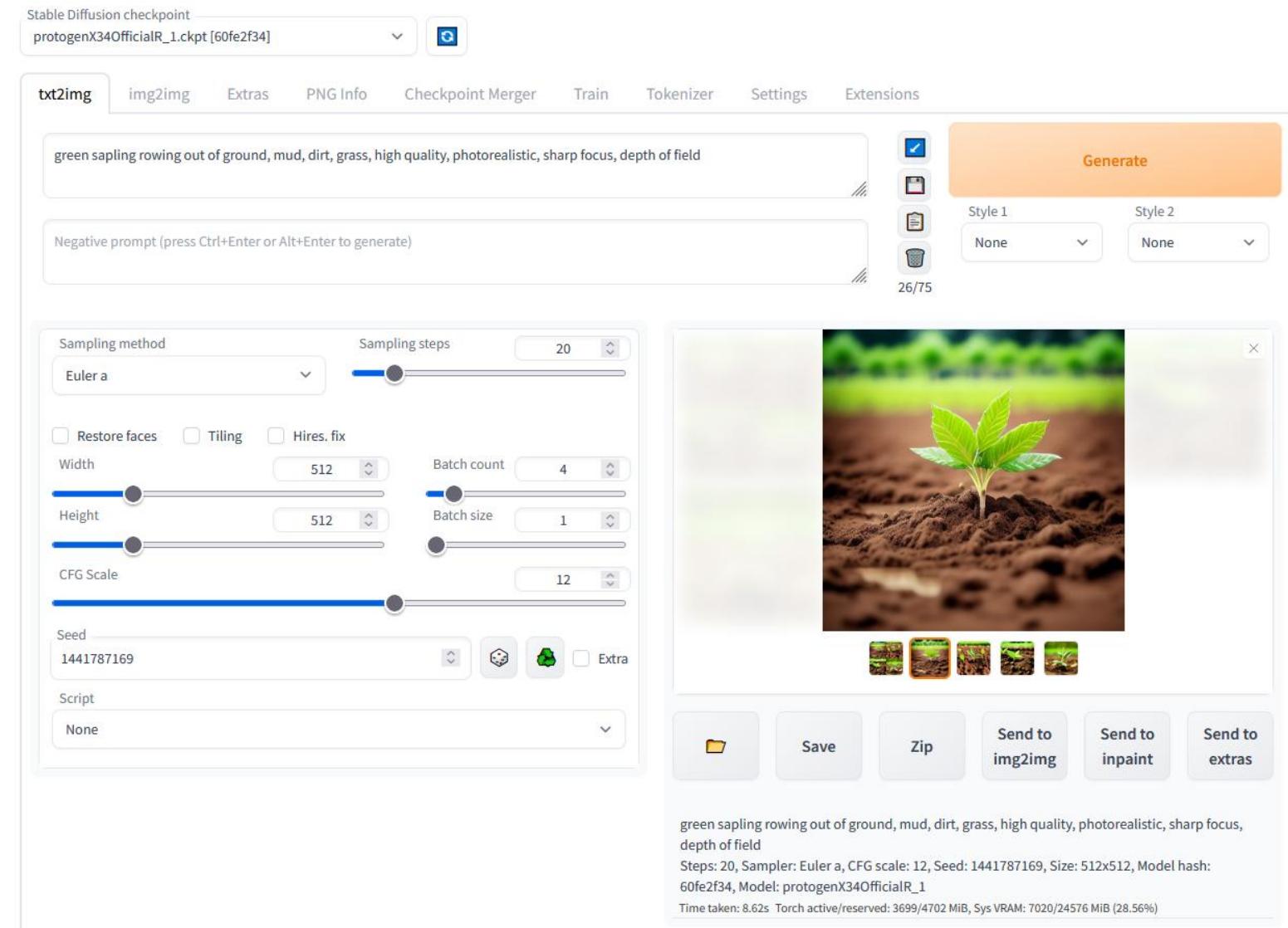
Installation on Windows 10/11 with NVidia-GPUs using release package

1. Download `sd.webui.zip` from [v1.0.0-pre](#) and extract its contents.
2. Run `update.bat`.
3. Run `run.bat`.

For more details see [Install-and-Run-on-NVidia-GPUs](#)

Automatic Installation on Windows

1. Install [Python 3.10.6](#) (Newer version of Python does not support torch), checking "Add Python to PATH".
2. Install [git](#).
3. Download the stable-diffusion-webui repository, for example by running `git clone https://github.com/AUTOMATIC1111/stable-diffusion-webui.git`
4. Run `webui-user.bat` from Windows Explorer as normal, non-administrator, user.



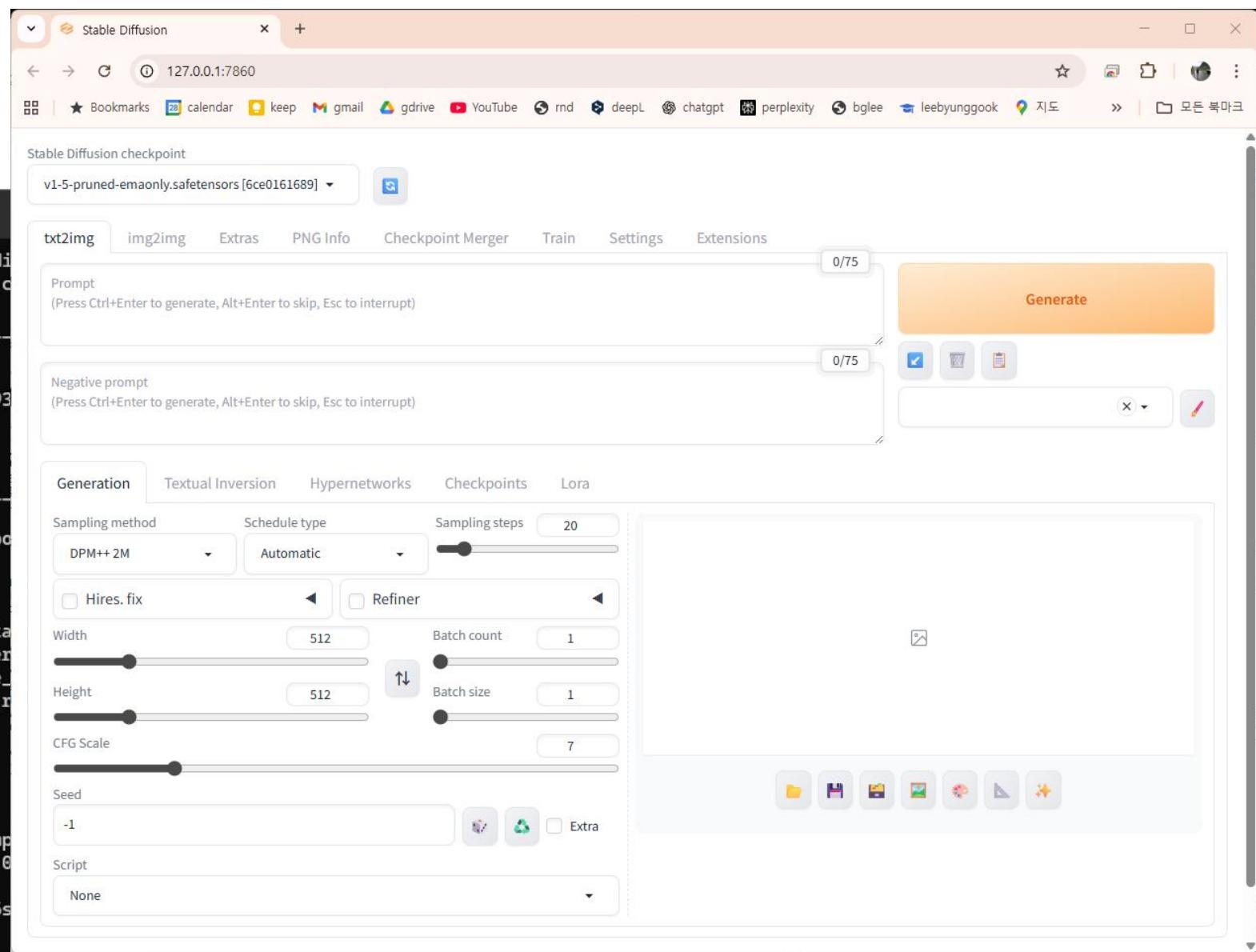
Stable Diffusion web UI A web interface for Stable Diffusion, implemented using Gradio library.

```
Windows PowerShell

The cache for model files in Transformers v4.22.0 has been updated. Migration. You can interrupt this and resume the migration later on by exit [00:00, ?it/s]
D:\stable-diffusion-webui\venv\lib\site-packages\timm\models\layers\models.layers is deprecated, please import via timm.layers
venv "D:\stable-diffusion-webui\venv\Scripts\Python.exe"
Python 3.10.8 (tags/v3.10.8:aaaf517, Oct 11 2022, 16:50:30) [MSC v.193
Version: v1.10.1
Commit hash: 82a973c04367123ae98bd9abdf80d9eda9b910e2
Launching Web UI with arguments:
D:\stable-diffusion-webui\venv\lib\site-packages\timm\models\layers\models.layers is deprecated, please import via timm.layers
    warnings.warn(f"Importing from {__name__} is deprecated, please import no module 'xformers'. Processing without...
no module 'xformers'. Processing without...
No module 'xformers'. Proceeding without it.
Loading weights [6ce0161689] from D:\stable-diffusion-webui\models\Sta
Creating model from config: D:\stable-diffusion-webui\configs\v1-infer
D:\stable-diffusion-webui\venv\lib\site-packages\huggingface_hub\file_
is deprecated and will be removed in version 1.0.0. Downloads always r
wnload, use 'force_download=True'.
    warnings.warn(
Running on local URL: http://127.0.0.1:7860

To create a public link, set `share=True` in `launch()`.

Startup time: 8.3s (prepare environment: 1.9s, import torch: 2.9s, imported: 0.2s, other imports: 0.4s, list SD models: 0.2s, load scripts: 0. Applying attention optimization: Doggett... done.
Model loaded in 2.3s (load weights from disk: 0.1s, create model: 0.5s, prompt: 0.1s).
```



ControlNet

2023.02

Adding Conditional Control to Text-to-Image Diffusion Models

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala
Stanford University

{lvmin, anyirao, maneesh}@cs.stanford.edu



Figure 1: Controlling Stable Diffusion with learned conditions. ControlNet allows users to add conditions like Canny edges (top), human pose (bottom), etc., to control the image generation of large pretrained diffusion models. The default results use the prompt “a high-quality, detailed, and professional image”. Users can optionally give prompts like the “chef in kitchen”.

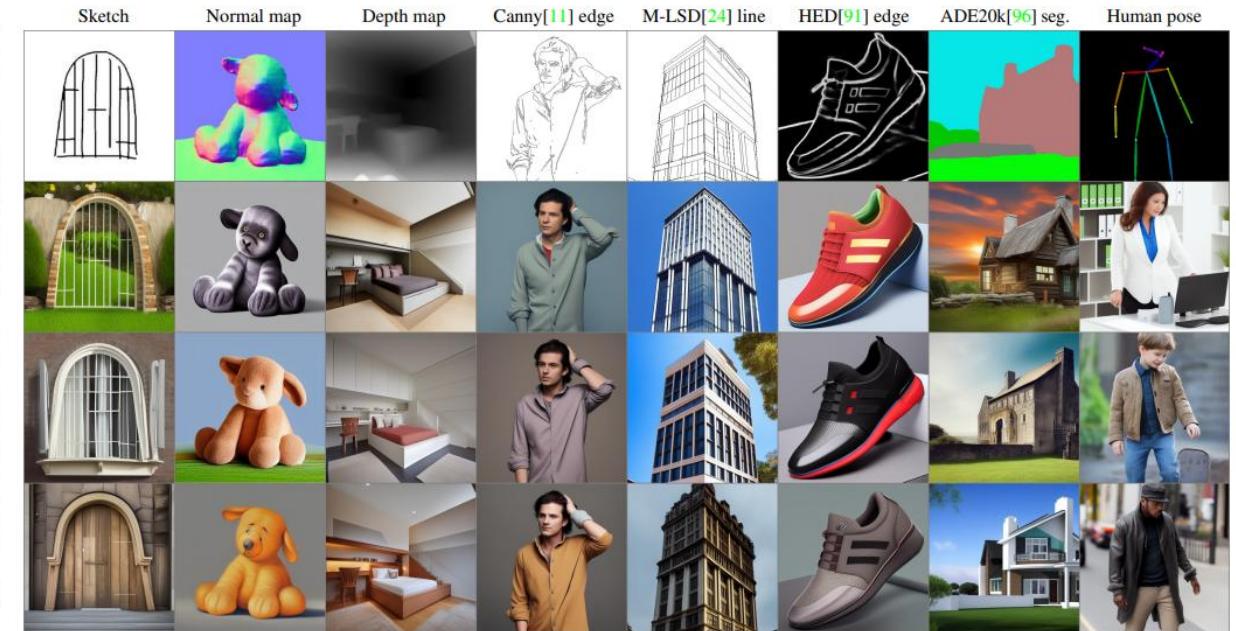


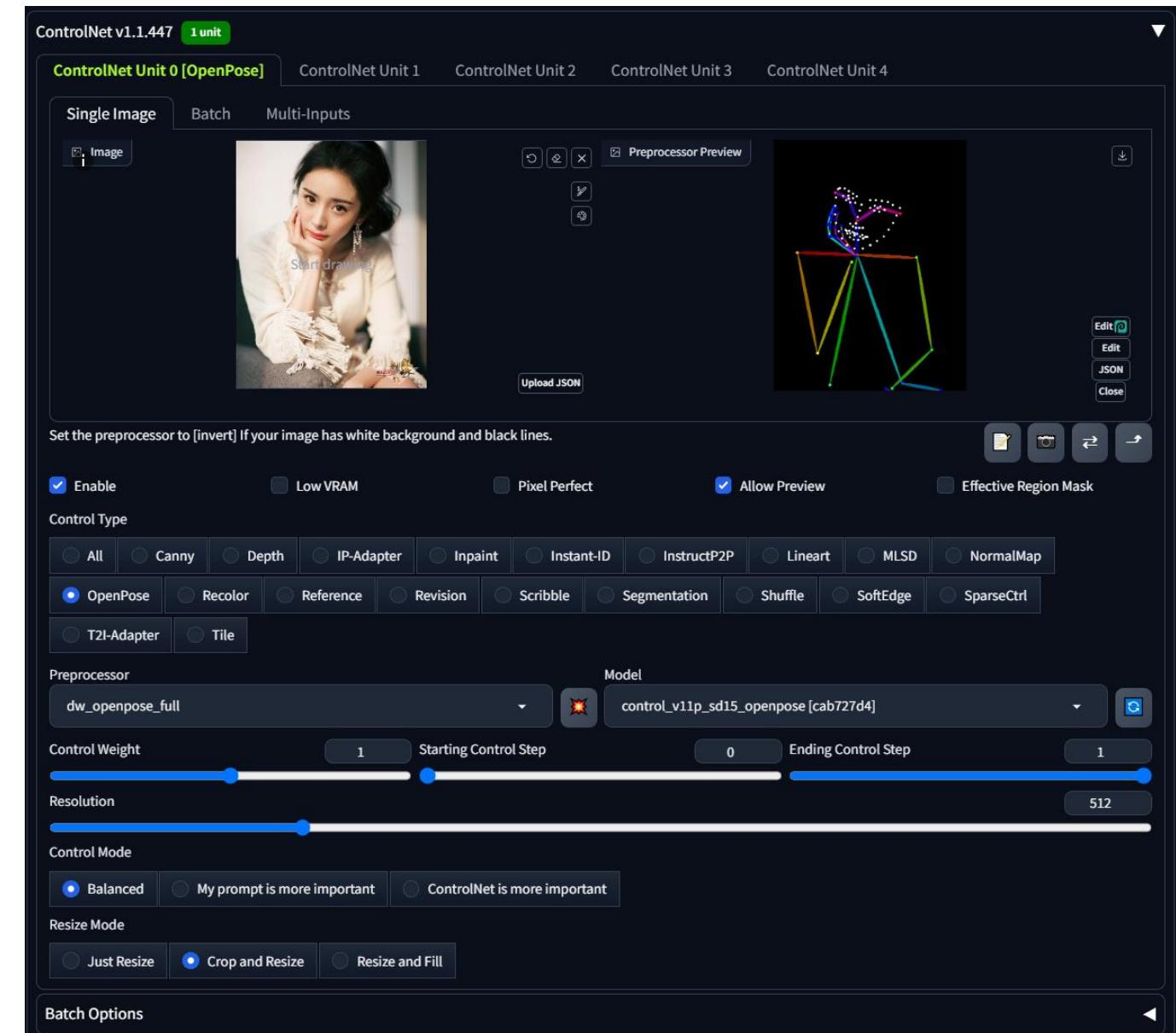
Figure 7: Controlling Stable Diffusion with various conditions **without prompts**. The top row is input conditions, while all other rows are outputs. We use the empty string as input prompts. All models are trained with general-domain data. The model has to recognize semantic contents in the input condition images to generate images.

ControlNet

ControlNet은 Stable Diffusion 생태계에서 아주 중요한 확장 기능 중 하나입니다. 간단히 말하면, 이미지 생성 과정에서 더 정교한 제어를 가능하게 해주는 모듈. 일반적인 Stable Diffusion은 프롬프트(텍스트)만으로 이미지를 생성하지만, ControlNet을 활용하면 추가 조건(조건 이미지)을 주어 모델이 원하는 구조, 포즈, 구도를 더 잘 따르도록 만들 수 있습니다.

ControlNet은 2023년 초 발표된 기술로, Stable Diffusion의 U-Net에 조건 입력(Condition Input)을 추가해 구조적 가이드를 따르면서도 창의적으로 이미지를 생성하도록

1. Open "Extensions" tab.
2. Open "Install from URL" tab in the tab.
3. Enter <https://github.com/Mikubill/sd-webui-controlnet.git> to "URL for extension's git repository".
4. Press "Install" button.
5. Wait for 5 seconds, and you will see the message "Installed into stable-diffusion-webui\extensions\ sd-webui-controlnet. Use Installed tab to restart".
6. Go to "Installed" tab, click "Check for updates", and then click "Apply and restart UI". (The next time you can also use these buttons to update ControlNet.)
7. Completely restart A1111 webui including your terminal. (If you do not know what is a "terminal", you can reboot your computer to achieve the same effect.)
8. Download models (see below).
9. After you put models in the correct folder, you may need to refresh to see the models. The refresh button is right to your "Model" dropdown.



The WebUI extension for ControlNet and other injection-based SD controls. This extension is for AUTOMATIC1111's [Stable Diffusion web UI](#), allows the Web UI to add [ControlNet](#) to the original54 Stable Diffusion model to generate images. The addition is on-the-fly, the merging is not required.

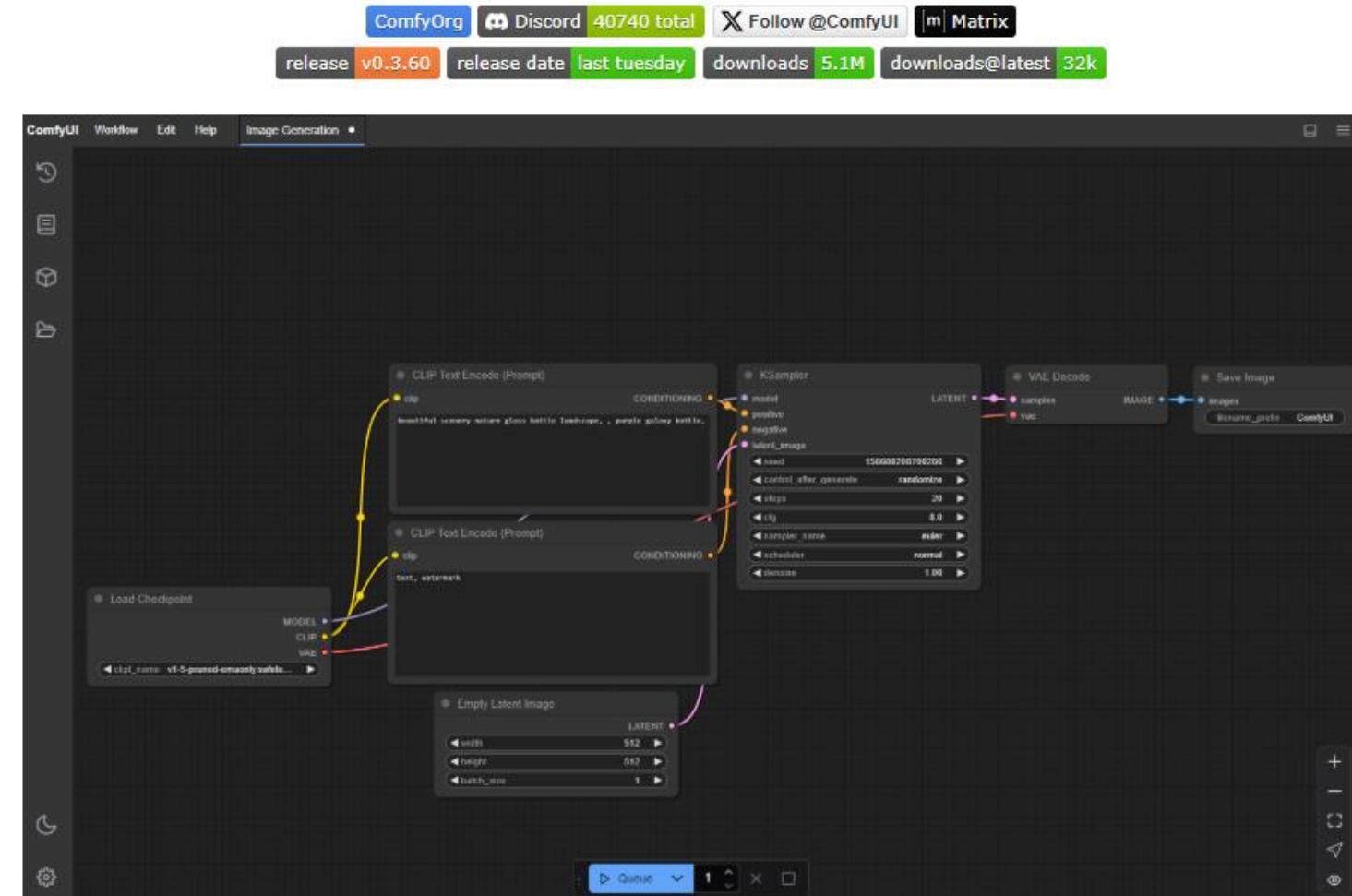
ComfyUI

2023.01

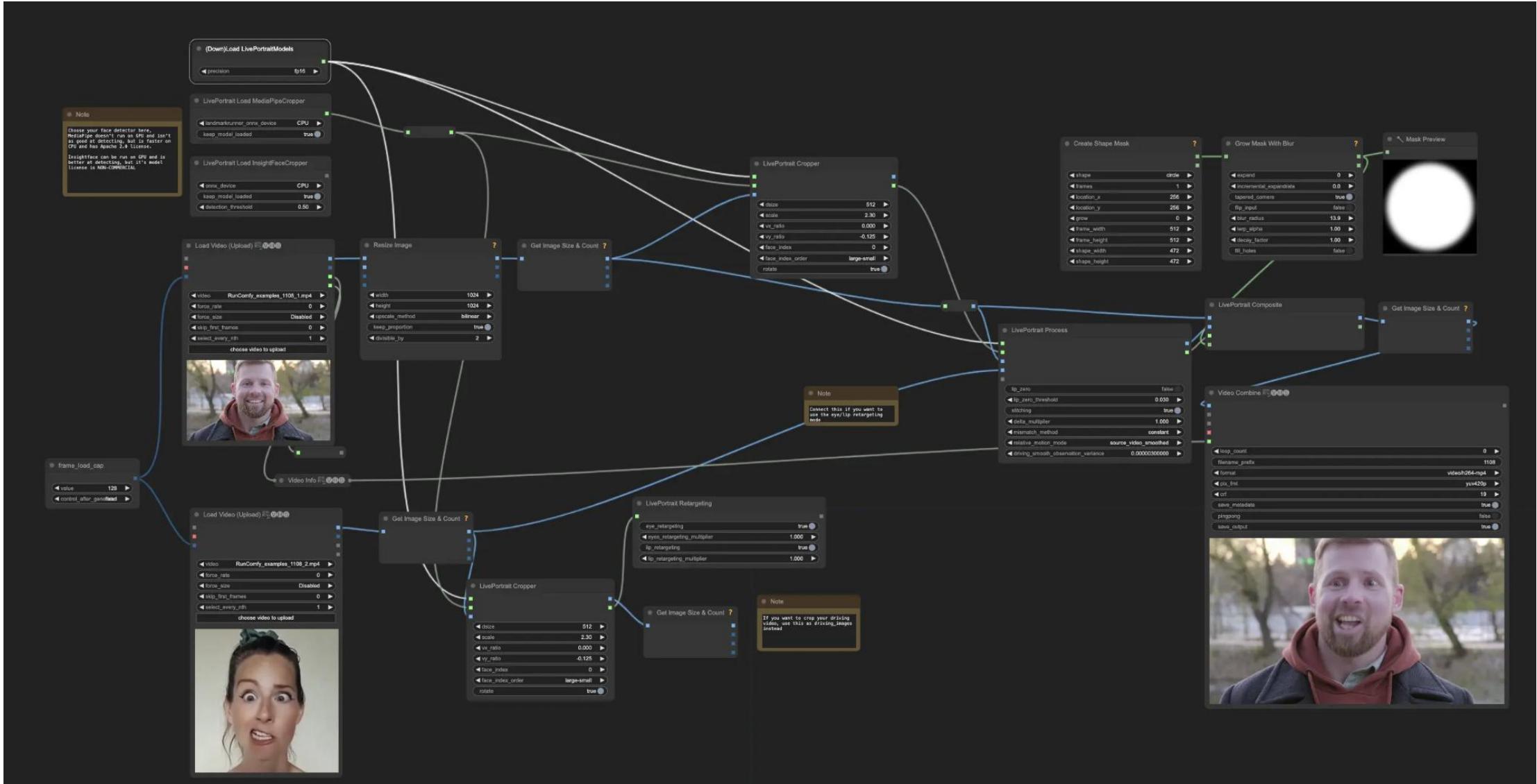
- Image Models
 - SD1.x, SD2.x ([unCLIP](#))
 - [SDXL](#), [SDXL Turbo](#)
 - [Stable Cascade](#)
 - [SD3 and SD3.5](#)
 - Pixart Alpha and Sigma
 - [AuraFlow](#)
 - [HunyuanDiT](#)
 - [Flux](#)
 - [Lumina Image 2.0](#)
 - [HiDream](#)
 - [Qwen Image](#)
 - [Hunyuan Image 2.1](#)
- Image Editing Models
 - [Omnigen 2](#)
 - [Flux Kontext](#)
 - [HiDream E1.1](#)
 - [Qwen Image Edit](#)
- Video Models
 - [Stable Video Diffusion](#)
 - [Mochi](#)
 - [LTX-Video](#)
 - [Hunyuan Video](#)
 - [Wan 2.1](#)
 - [Wan 2.2](#)
- Audio Models
 - [Stable Audio](#)
 - [ACE Step](#)
- 3D Models
 - [Hunyuan3D 2.0](#)

ComfyUI

The most powerful and modular visual AI engine and application.



ComfyUI lets you design and execute advanced stable diffusion pipelines using a graph/nodes/flowchart based interface. Available on Windows, Linux, and macOS.



ComfyUI is an [open source](#), [node-based](#) program that allows users to generate images from a series of [text prompts](#). It uses free [diffusion models](#) such as [Stable Diffusion](#) as the base model for its image capabilities combined with other tools such as [ControlNet](#) and [LCM Low-rank adaptation](#) with each tool being represented by a node in the program. ComfyUI는 Stable Diffusion 모델을 위한 강력하고 모듈식의 GUI(Graphical User Interface) 및 백엔드 시스템입니다. 이 인터페이스는 그래프, 노드, 플로우차트 기반의 인터페이스를 사용하여 복잡한 Stable Diffusion 작업 흐름을 설계하고 실행할 수 있게 해주는 것이 특징. 사용자는 코드를 작성할 필요 없이 Stable Diffusion 작업 흐름을 실험하고 생성할 수 있습니다.

<https://www.anaconda.com/> <https://www.anaconda.com/products/individual> <https://www.nvidia.com/en-us/drivers/>
<https://git-scm.com/downloads/win> <https://github.com/comfyanonymous/ComfyUI.git>



ComfyUI

install git
Windows Terminal

```
git clone https://github.com/comfyanonymous/ComfyUI.git  
cd ComfyUI
```

```
install anaconda  
Anaconda Powershell Prompt(Anaconda3)  
conda update conda  
conda create -n ComfyUI python=3.12 -y  
conda activate ComfyUI
```

```
cd ComfyUI  
conda install pytorch torchvision torchaudio pytorch-cuda=12.1 -c pytorch -c nvidia -y  
pip install -r requirements.txt
```

CondaSSLError: OpenSSL appears to be unavailable on this machine.

OpenSSL 모듈을 찾지 못하여 발생하는 오류로, anaconda3\Library\bin 폴더

[libcrypto-1_1-x64.dll](#) 와 [libssl-1_1-x64.dll](#) 파일을 복사

anaconda3\Dlls 에 붙여 넣고 다시 시도

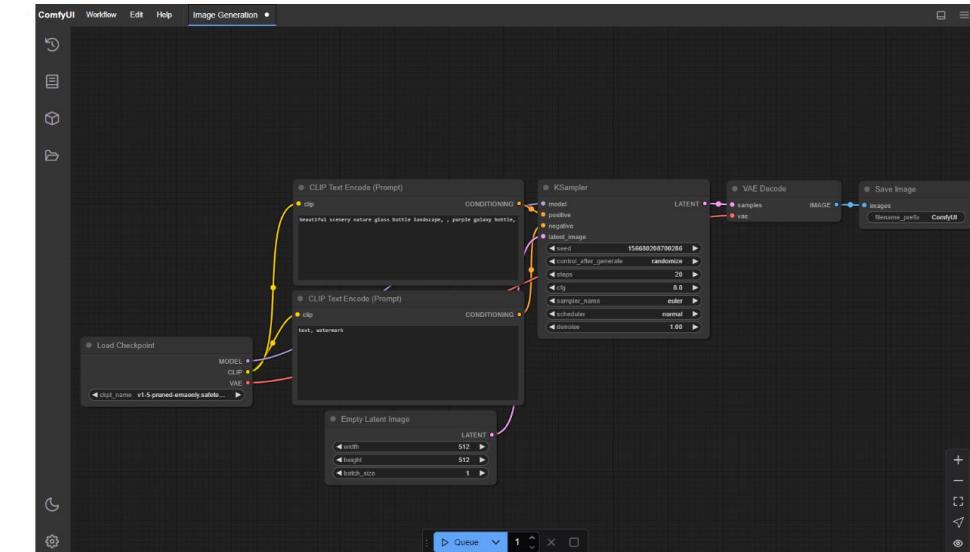
<https://wolfzone.tistory.com/53>



ComfyUI

The most powerful and modular visual AI engine and application.

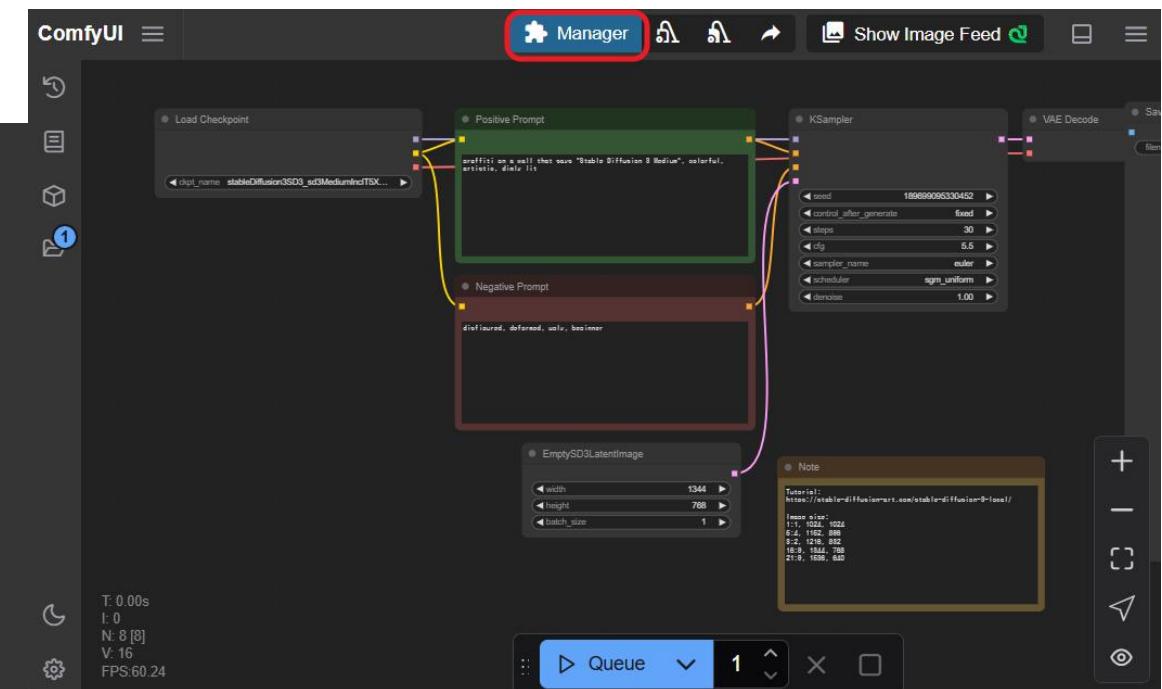
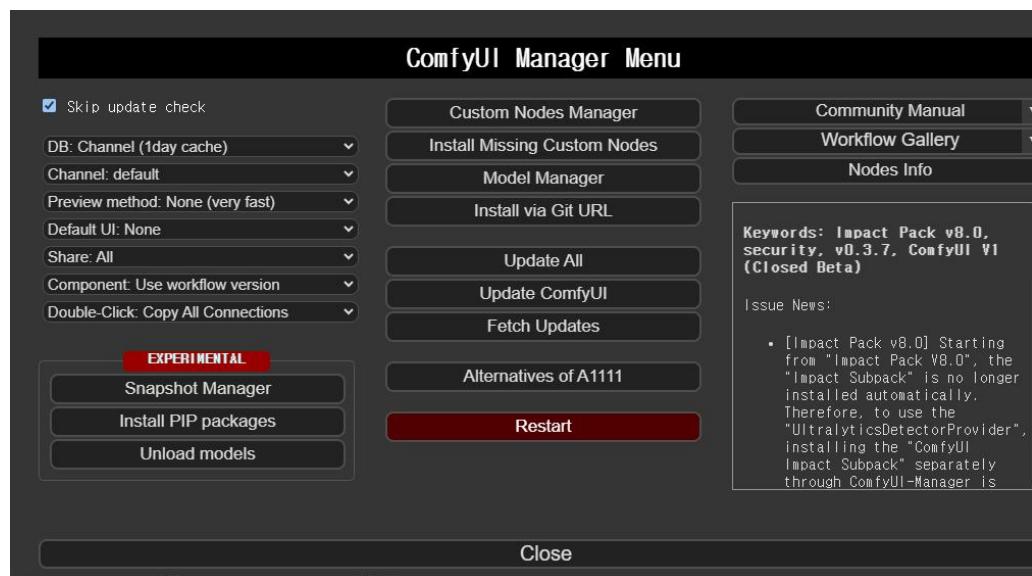
ComfyOrg Discord 39856 total X Follow @ComfyUI Matrix
release v0.3.59 release date september downloads 5M downloads@latest 39k



ComfyUI lets you design and execute advanced stable diffusion pipelines using a graph/nodes/flowchart based interface. Available on Windows, Linux, and macOS.

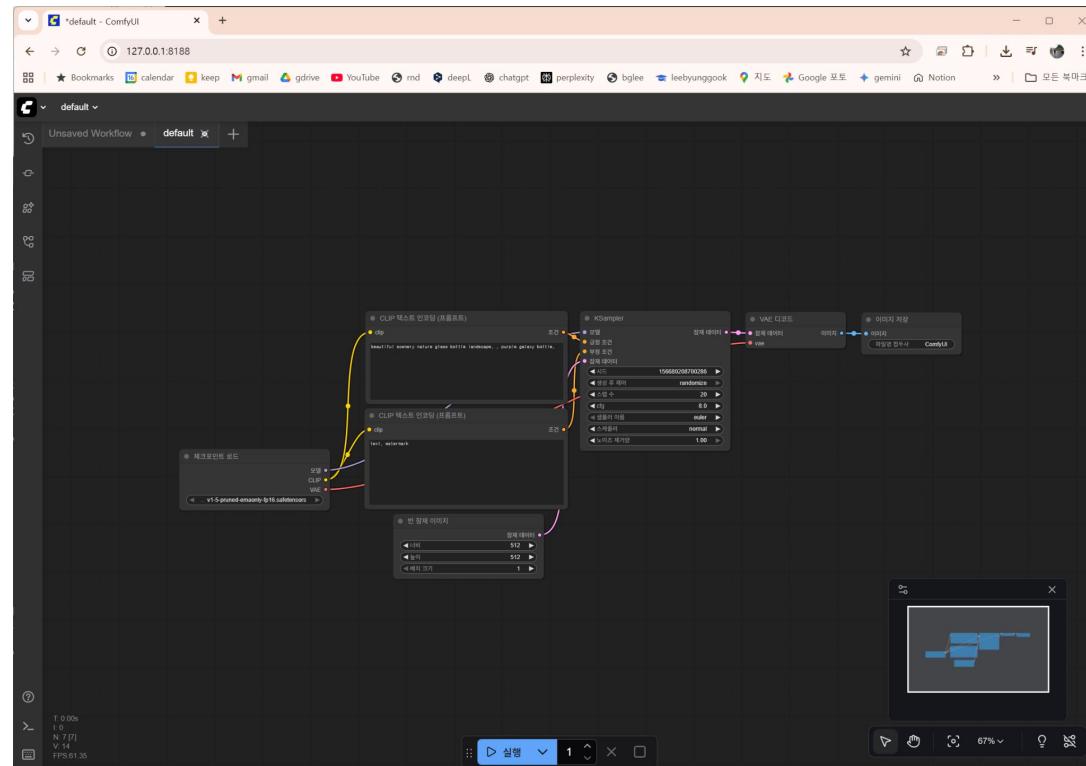
ComfyUI manager

```
cd \ComfyUI\custom_nodes  
git clone https://github.com/ltdrdata/ComfyUI-Manager
```



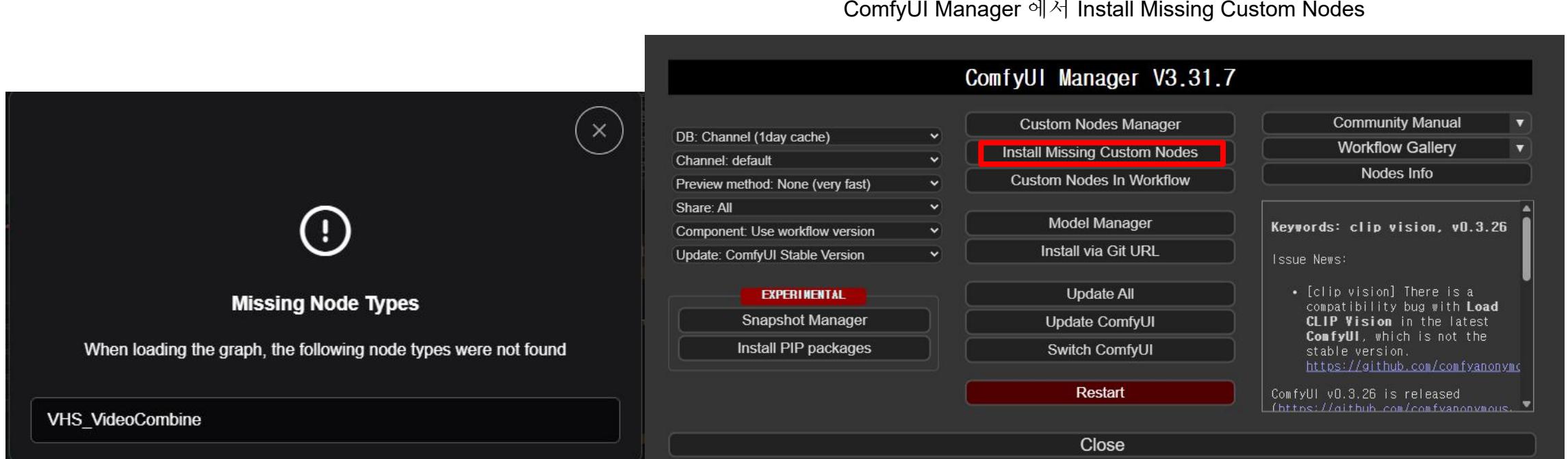
ComfyUI

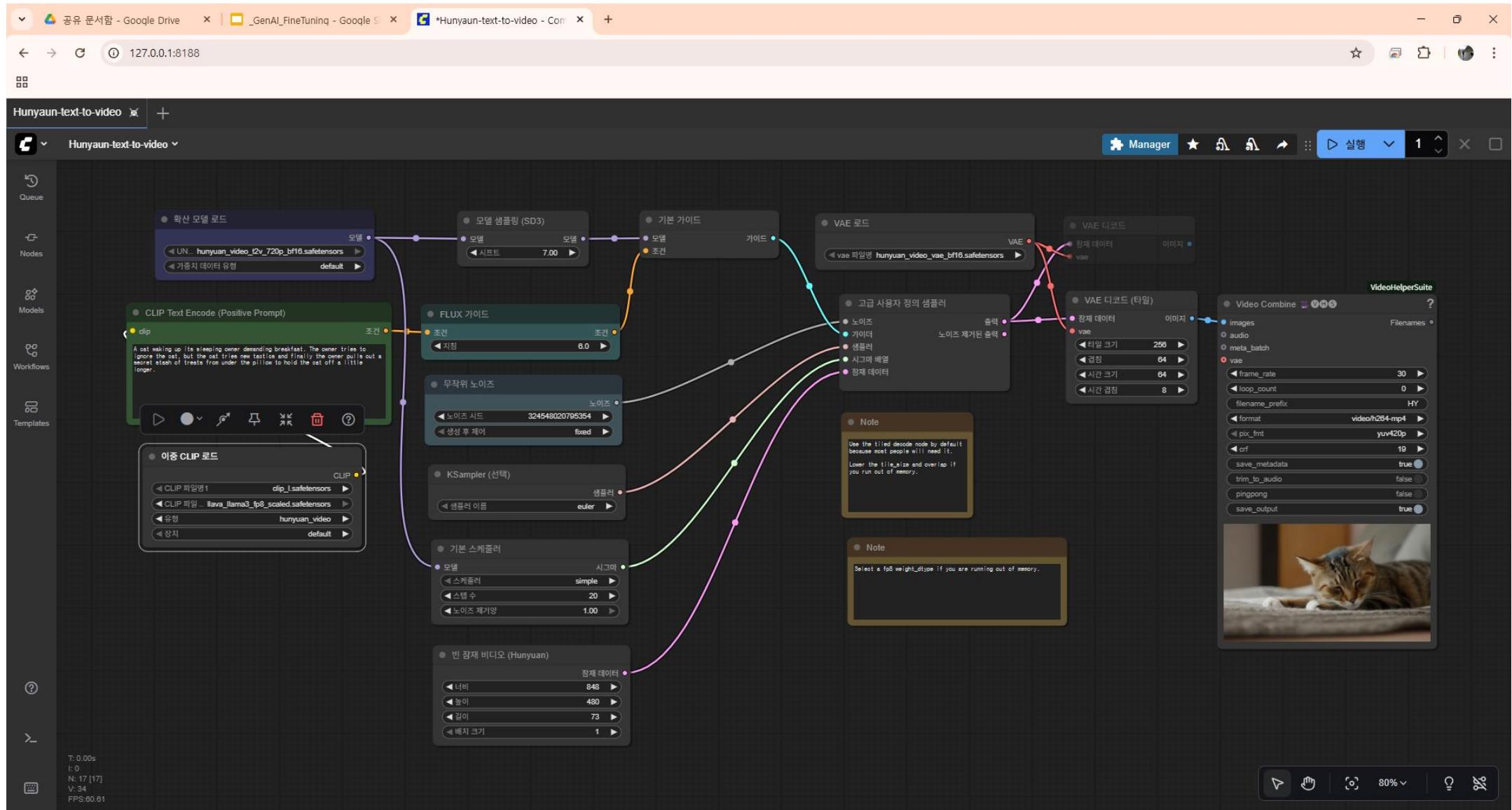
```
conda activate ComfyUI  
cd ComfyUI  
python main.py  
open Hunyaun-text-to-video.json
```



```
Anaconda Powershell Prompt  
Downloaded regex-2025.9.18-cp312-cp312-win_amd64.whl (275 kB)  
Using cached trampoline-0.1.2-py3-none-any.whl (5.2 kB)  
Using cached typing-inspection-0.4.1-py3-none-any.whl (14 kB)  
Using cached colorama-0.4.6-py2.py3-none-any.whl (25 kB)  
Using cached mako-1.3.10-py3-none-any.whl (78 kB)  
Installing collected packages: trampoline, typing-inspection, sympy, sentencepiece, scipy, safetensors, regex, python-dotenv, pydantic-core, psutil, propcache, packaging, multidict, Mako, kornia_rs, greenlet, fsspec, frozenlist, einops, comfyui-workflow-templates, comfyui-frontend-package, comfyui-embedded-docs, colorama, av, attrs, annotated-types, aiohappyeyeballs, yarl, tqdm, SQLAlchemy, soundfile, pydantic, aiosignal, torchsde, pydantic-settings, kornia, huggingface-hub, alembic, aiohttp, tokenizers, spandrel, transformers  
Attempting uninstall: sympy  
  Found existing installation: sympy 1.14.0  
    Uninstalling sympy-1.14.0:  
      Successfully uninstalled sympy-1.14.0  
Successfully installed Mako-1.3.10 SQLAlchemy-2.0.43 aiohappyeyeballs-2.6.1 aiohttp-3.12.15 aiosignal-1.4.0 alembic-1.16.5 annotated-types-0.7.0 attrs-25.3.0 av-15.1.0 colorama-0.4.6 comfyui-embedded-docs-0.2.6 comfyui-frontend-package-1.26.11 comfyui-workflow-templates-0.1.81 einops-0.8.1 frozenlist-1.7.0 fsspec-2025.9.0 greenlet-3.2.4 huggingface-hub-0.35.0 kornia-0.8.1 kornia_rs-0.1.9 multidict-6.6.4 packaging-25.0 propcache-0.3.2 psutil-7.1.0 pydantic-2.11.9 pydantic-core-2.33.2 pydantic-settings-2.10.1 python-dotenv-1.1.1 regex-2025.9.18 safetensors-0.6.2 scipy-1.16.2 sentencepiece-0.2.1 soundfile-0.13.1 spandrel-0.4.1 sympy-1.13.1 tokenizers-0.22.0 torchsde-0.2.6 tqdm-4.67.1 trampoline-0.1.2 transformers-4.56.1 typing-inspection-0.4.1 yarl-1.20.1  
(ComfyUI) PS D:\Hunyaun\ComfyUI> python main.py  
Checkpoint files will always be loaded safely.  
Total VRAM 8192 MB, total RAM 65293 MB  
pytorch version: 2.5.1  
Set vram state to: NORMAL_VRAM  
Device: cuda:0 NVIDIA GeForce RTX 3060 Ti : native  
Torch version too old to set sdpa backend priority.  
Using pytorch attention  
Python version: 3.12.11 | packaged by Anaconda, Inc. | (main, Jun 5 2025, 12:58:53) [MSC v.1929 64 bit (AMD64)]  
ComfyUI version: 0.3.59  
***** User settings have been changed to be stored on the server instead of browser storage. *****  
***** For multi-user setups add the --multi-user CLI argument to enable multiple user profiles. *****  
ComfyUI frontend version: 1.26.11  
[Prompt Server] web root: C:\Users\HP\conda\envs\ComfyUI\Lib\site-packages\comfyui_frontend_package\static  
Import times for custom nodes:  
  0.0 seconds: D:\Hunyaun\ComfyUI\custom_nodes\websocket_image_save.py  
Context impl SQLiteImpl.  
Will assume non-transactional DDL.  
No target revision found.  
Starting server  
To see the GUI go to: http://127.0.0.1:8188
```

Error





<https://www.internetmap.kr/entry/Stable-Diffusion-via-ComfyUI> <https://www.internetmap.kr/entry/Installing-ComfyUI-Desktop>

<https://www.internetmap.kr/entry/Mochi-GGUF-txt2vid-ComfyUI>

ComfyUI

기본적으로 ComfyUI는 PC뿐만 아니라, Mac, Linux에서도 실행시킬 수 있습니다.

Google Colab과 같은 클라우드 환경에서도 실행시킬 수 있습니다.

대략 VRAM 8GB, GTX3060 이상의 GPU를 사용하는 것이 좋습니다.

RTX3070 VRAM 8GB에서 SDXL 모델로 이미지를 생성할 경우 10~15초 정도면 이미지를 생성할 수 있고,
FLUX 모델을 사용할 경우(원래는 20GB가 최소사양이지만, [GGUF 모델을 사용할 경우](#)) 1분 10초 정도면 이미지를 생성할 수 있습니다.

[LTX](#), [CogVideo](#), [GGUF](#), [Mochi GGUF](#) 등 비디오 모델의 경우, 50프레임 정도를 기준으로 할 때 대략 10분~20분 정도면 생성할 수 있다.

무엇보다 ComfyUI 데스크톱은 WebUI 버전에 비해 자원을 덜 소모하고 생성속도도 빠른 장점이 있다.

ComfyUI 데스크톱은 오픈 소스 프로젝트로서, 필요하시면 [여기](#)에서 전체 코드를 다운로드 받으실 수 있다.

ComfyUI를 실행시키려면 윈도 10 이상과 NVIDIA GPU 가 필요합니다.



LTX-Video

1. <https://huggingface.co/Lightricks/LTX-Video/blob/main/ltx-video-2b-v0.9.safetensors>

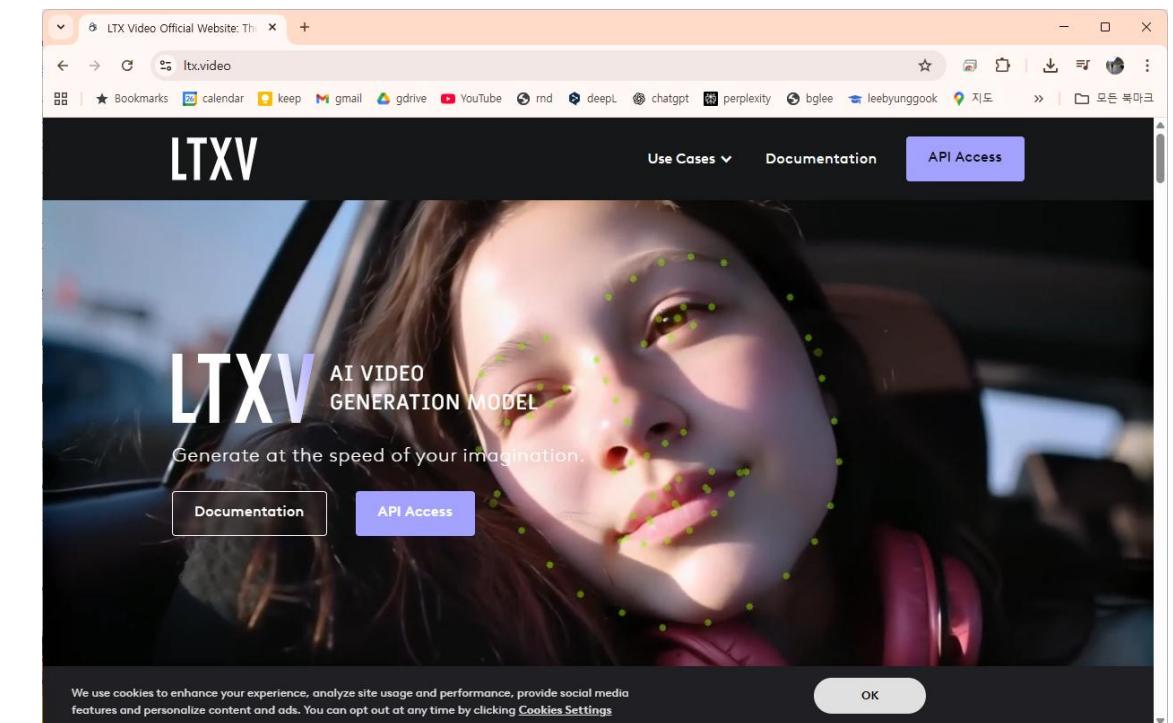
ComfyUI\models\checkpoints Size of remote file:9.37GB

2. https://huggingface.co/Comfy-Org/stable-diffusion-3.5-fp8/blob/main/text_encoders/t5xxl_fp16.safetensors

ComfyUI\models\text_encoders Size of remote file:9.79GB

3. [LTX-img2vid-1.json](#)

ComfyUI Workflow->Open JSON 파일



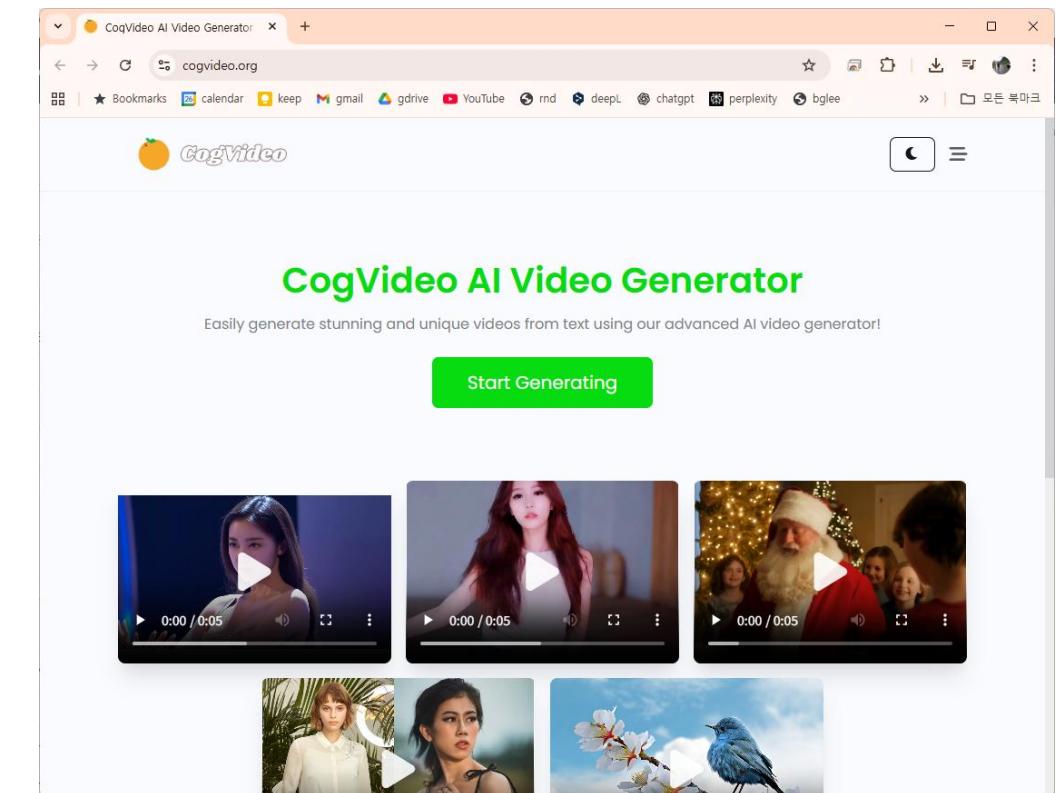
CogVideo

1. https://huggingface.co/comfyanonymous/flux_text_encoders/blob/main/t5xxl_fp8_e4m3fn.safetensors

ComfyUI\models\clip Size of remote file:4.89GB

2. [cogvideox_i2v_GGUF.json](#)

ComfyUI Workflow->Open JSON 파일



Mochi GGUF

1. <https://huggingface.co/calcuis/pig/tree/main>

pig-mochi...gguf 파일을 ComfyUI\models\diffusion_models

2. https://huggingface.co/city96/t5-v1_1-xxl-encoder-gguf/tree/main

다운로드 받은 unet과 동일한 버전 파일을 ComfyUI\models\clip

3. https://huggingface.co/Comfy-Org/mochi_preview_repackaged/tree/main/split_files/vae

ComfyUI\models\vae Size of remote file:920MB

4. [Mochi_GGUF_movie_4clips.json](#)

ComfyUI Workflow->Open JSON 파일

run it with gguf-node via comfyui

- drag **mochi** to > ./ComfyUI/models/diffusion_models
- drag **t5xxl** to > ./ComfyUI/models/text_encoders
- drag **vae** to > ./ComfyUI/models/vae
- drag demo video (below) to > your browser for workflow



Prompt ⓘ
a pinky pig moving quickly in a
beautiful winter scenery nature
trees sunset tracking camera



Prompt ⓘ
a pinky pig moving quickly in a
beautiful winter scenery nature
trees sunset tracking camera



Prompt ⓘ
a pinky pig moving quickly in a
beautiful winter scenery nature
trees sunset tracking camera

HunyuanVideo and World

HunyuanVideo: A Systematic Framework For Large Video Generation Model

HunyuanVideo is a large-scale open-source video generation foundation model with 13 billion parameters based on text-to-video, released alongside a paper titled “A Systematic Framework For Large Video Generative Models.” It often outperforms existing closed models such as Runway Gen-3 and Luma 1.6 in terms of quality.

| Model | Setting (height/width/frame) | GPU Peak Memory |
|--------------|---------------------------------|-----------------|
| HunyuanVideo | 720px1280px129f | 60GB |
| HunyuanVideo | 544px960px129f | 45GB |

- An NVIDIA GPU with CUDA support is required.
 - The model is tested on a single 80G GPU.
 - **Minimum:** The minimum GPU memory required is 60GB for 720px1280px129f and 45G for 544px960px129f.
 - **Recommended:** We recommend using a GPU with 80GB of memory for better generation quality.

HunyuanVideo를 사용하여 848x480(480p) 73 프레임을 생성하는데 걸리는 시간
RTX3070에서 13.8분 RTX4090에서 4.5분 Google Colab(L4 런타임)에서 11분
하드웨어 요구사항이 워크플로를 돌리려면 NVidia GPU카드가 필요
최소 8GB VRAM에서도 Hunyuan Video를 돌릴 수 있다.

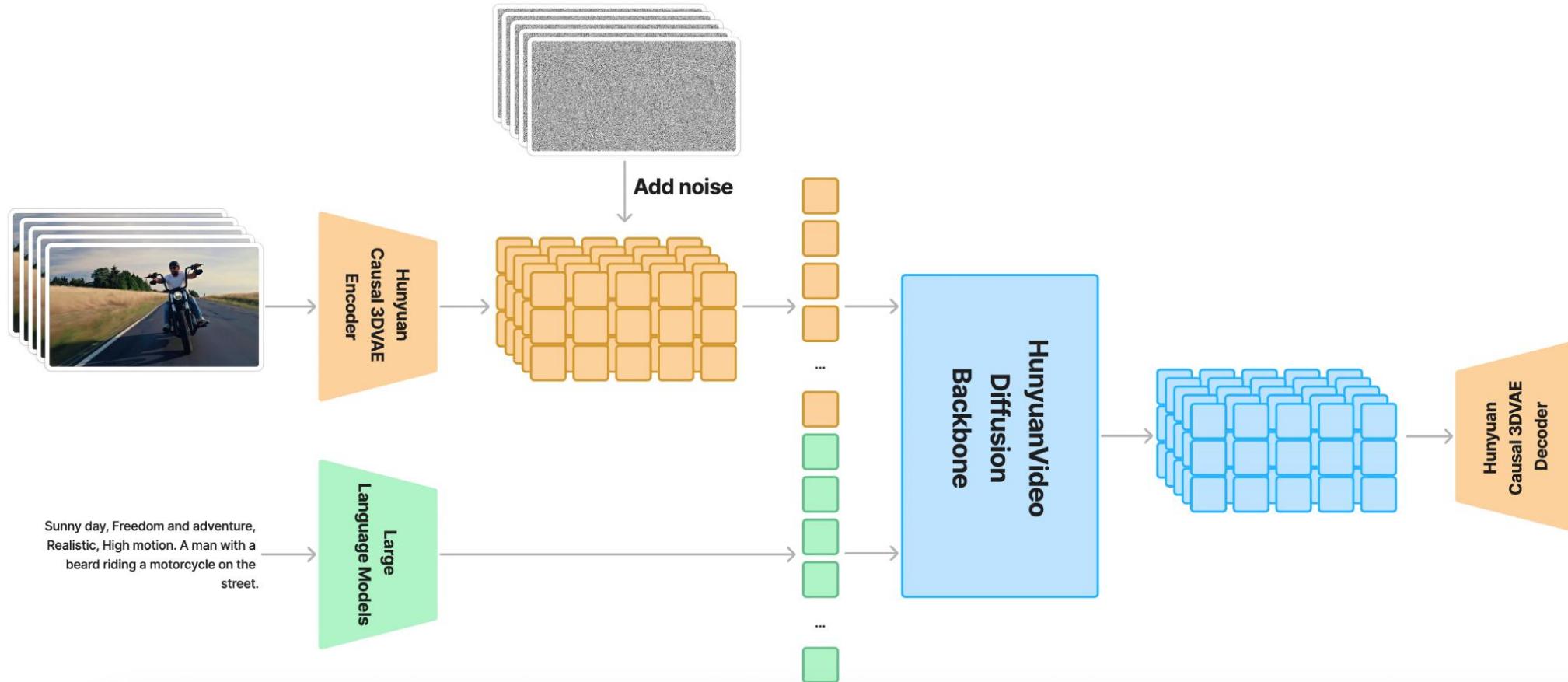
2024.12



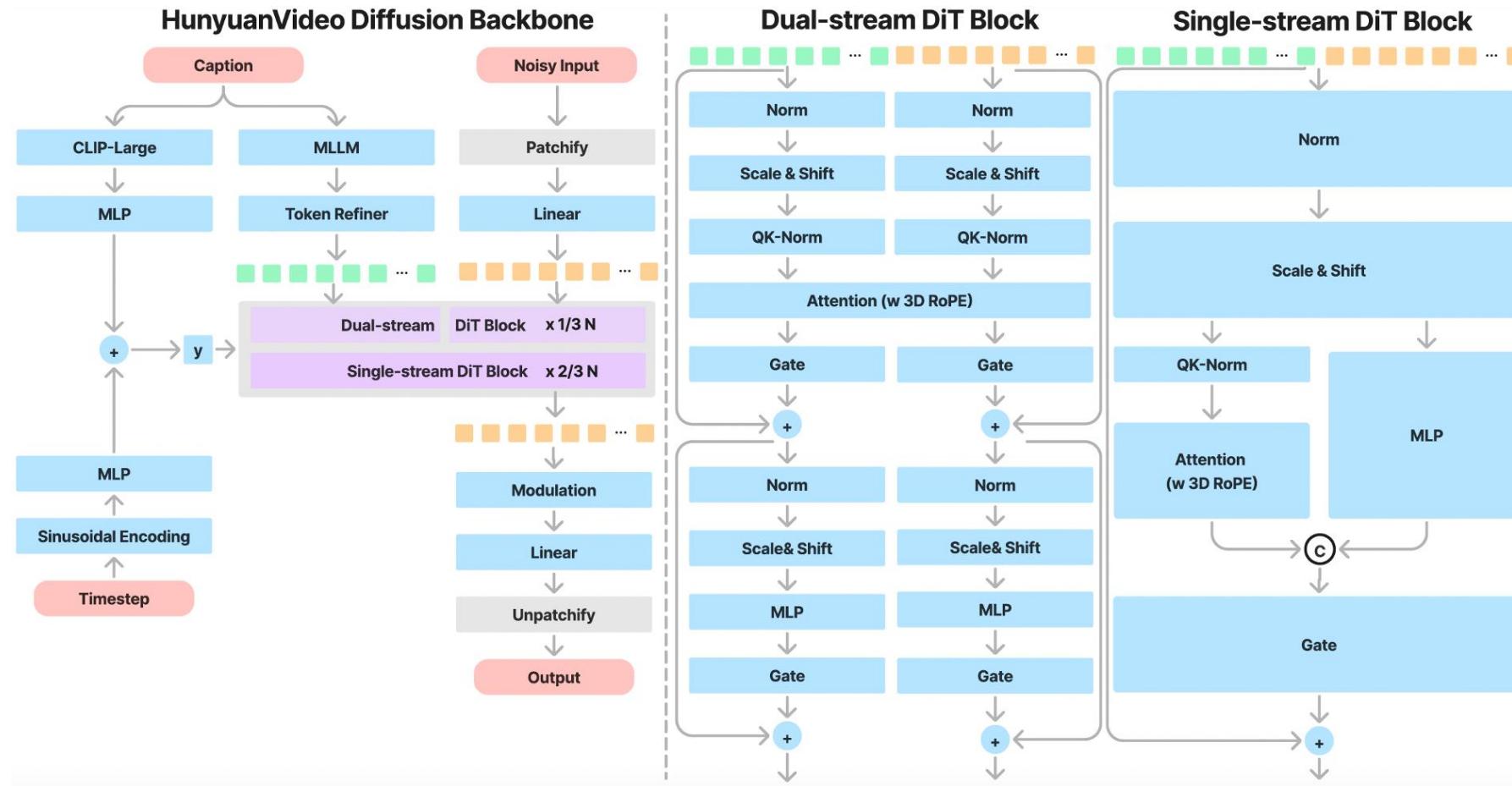


HunyuanVideo

2024.12



HunyuanVideo



HunyuanVideo(T2V)

1. split_files/diffusion_models/hunyuan_video_t2v_720p_bf16.safetensors

https://huggingface.co/Comfy-Org/HunyuanVideo_repackaged/blob/main/split_files/diffusion_models/hunyuan_video_t2v_720p_bf16.safetensors

ComfyUI\models\diffusion_models Size of remote file:25.6GB

2. split_files/text_encoders/llava_llama3_fp8_scaled.safetensors

https://huggingface.co/Comfy-Org/HunyuanVideo_repackaged/blob/main/split_files/text_encoders/clip_l.safetensors

ComfyUI\models\text_encoders Size of remote file:246MB

https://huggingface.co/Comfy-Org/HunyuanVideo_repackaged/blob/main/split_files/text_encoders/llava_llama3_fp8_scaled.safetensors

ComfyUI\models\text_encoders Size of remote file:9.09GB

3. split_files/vae/hunyuan_video_vae_bf16.safetensors

https://huggingface.co/Comfy-Org/HunyuanVideo_repackaged/blob/main/split_files/vae/hunyuan_video_vae_bf16.safetensors

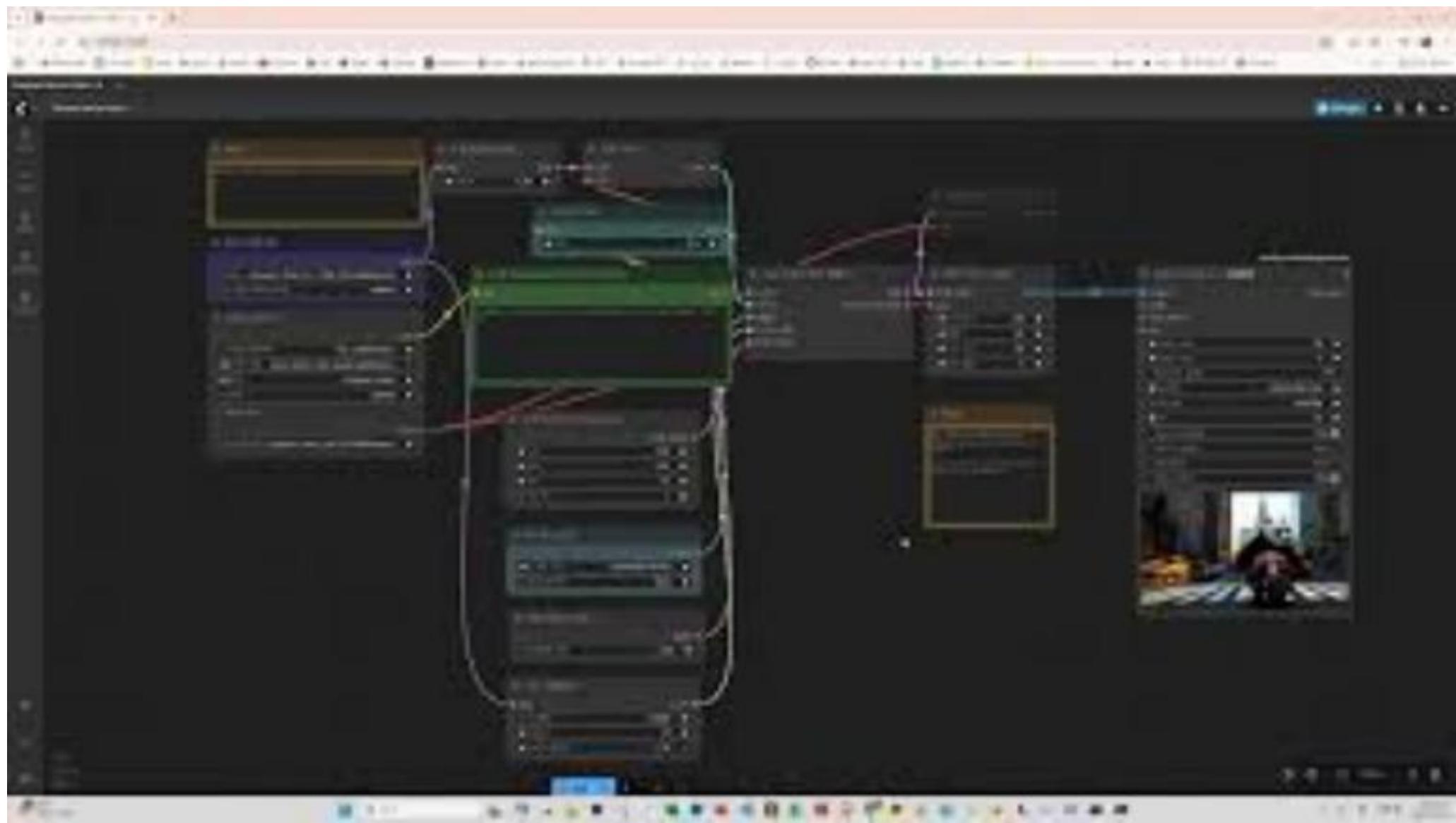
ComfyUI\models\vae Size of remote file:493MB

4. Hunyaun-text-to-video.json

<https://blog.kakaocdn.net/dn/czCJMX/btsLQ70XxJh/iE44mhr9zKTKFOyK1DUwr0/Hunyaun-text-to-video.json?attach=1&knm=tfile.json>

ComfyUI Workflow->Open JSON 파일

HW5



ComfyUI

ComfyUI 아카데미 소개

- 레슨 1 : ComfyUI 사용하기 기본(Using ComfyUI, EASY basics)
 - ComfyUI Manager 설치
 - 워크플로 생성시 참고사항
 - Reroute 노드 활용
 - 입력슬롯과 입력위젯 바꾸기
- 레슨 2 : 멋진 ComfyUI txt2img 방법
 - 비슷한 이미지 동시 생성
 - ControlNet을 사용한 개선
 - 최종 워크플로
- 레슨 3 : 잠재 이미지 확대(Latent Upscaling)
- 레슨 4 : Image to Image 생성
 - 기본 img2img
 - 고급 img2img
- 레슨 5 : 마술 같은 Img2Img 와 WD14
 - 기본 Img2Img
 - 프롬프트가 없는 Img2Img
- 레슨 6 : 모델 바꿔치기 및 마스킹
 - 모델 바꿔치기
 - 이미지 마스크 사용법
- 레슨 7 : 실시간 모델 합치기(Merge)
 - AUTOMATIC1111와 비교
- 레슨 8 : ComfyUI에서 LoRA 사용법
 - LoRA란?
 - 기본적인 LoRA 사용법
 - LoRA STACK
- 레슨 9 : ComfyUI에서 ControlNet 사용법
 - 일반적인 ControlNet 사용법
 - 전처리기(Preprocessor)
- 레슨 10 : UltimateSDUpscaler
 - Upscaler 가 필요한 이유
 - Upscaler 워크플로
 - Upscale 모델
- 레슨 11 : Face Detailer
 - Face Detailer 가 필요한 이유
 - Face Detailer 워크플로
 - 입력 이미지에 노이즈 추가
 - 얼굴 복원

Stable Diffusion

1. 개요

- Stable Diffusion은 2022년 Stability AI가 공개한 **Latent Diffusion Model (LDM)** 기반의 오픈소스 이미지 생성 모델
- 텍스트 프롬프트를 입력하면 해당 설명에 맞는 이미지를 생성할 수 있음.
- 누구나 PC에서 실행할 수 있을 정도로 경량화되어 있어 Midjourney, DALL·E 같은 클라우드 기반 모델과 달리 자유롭게 커스터마이즈 가능.

2. 핵심 아이디어 – Latent Diffusion

기존 Diffusion Model은 픽셀 공간에서 직접 노이즈를 제거했기 때문에 계산량이 매우 컸다. SD는 다음과 같은 최적화를 적용했음

- VAE(Variational Autoencoder)를 사용하여 이미지를 잠재 공간(latent space)으로 인코딩
- 잠재 공간에서 Diffusion을 수행 → 연산량과 메모리 사용량 크게 감소
- 최종 단계에서 다시 디코딩해 원본 이미지로 변환 즉, Stable Diffusion은 "압축된 이미지 표현" 위에서 Diffusion을 수행하므로 빠르고 효율적

3. 아키텍처 구성

Stable Diffusion은 크게 세 부분으로 나눌 수 있다.

1. Text Encoder

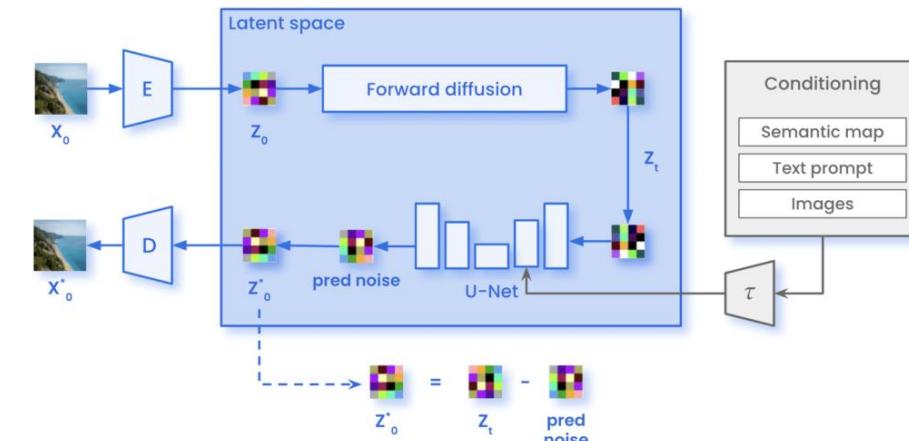
- CLIP의 텍스트 인코더를 사용
- 텍스트 프롬프트를 임베딩 벡터로 변환

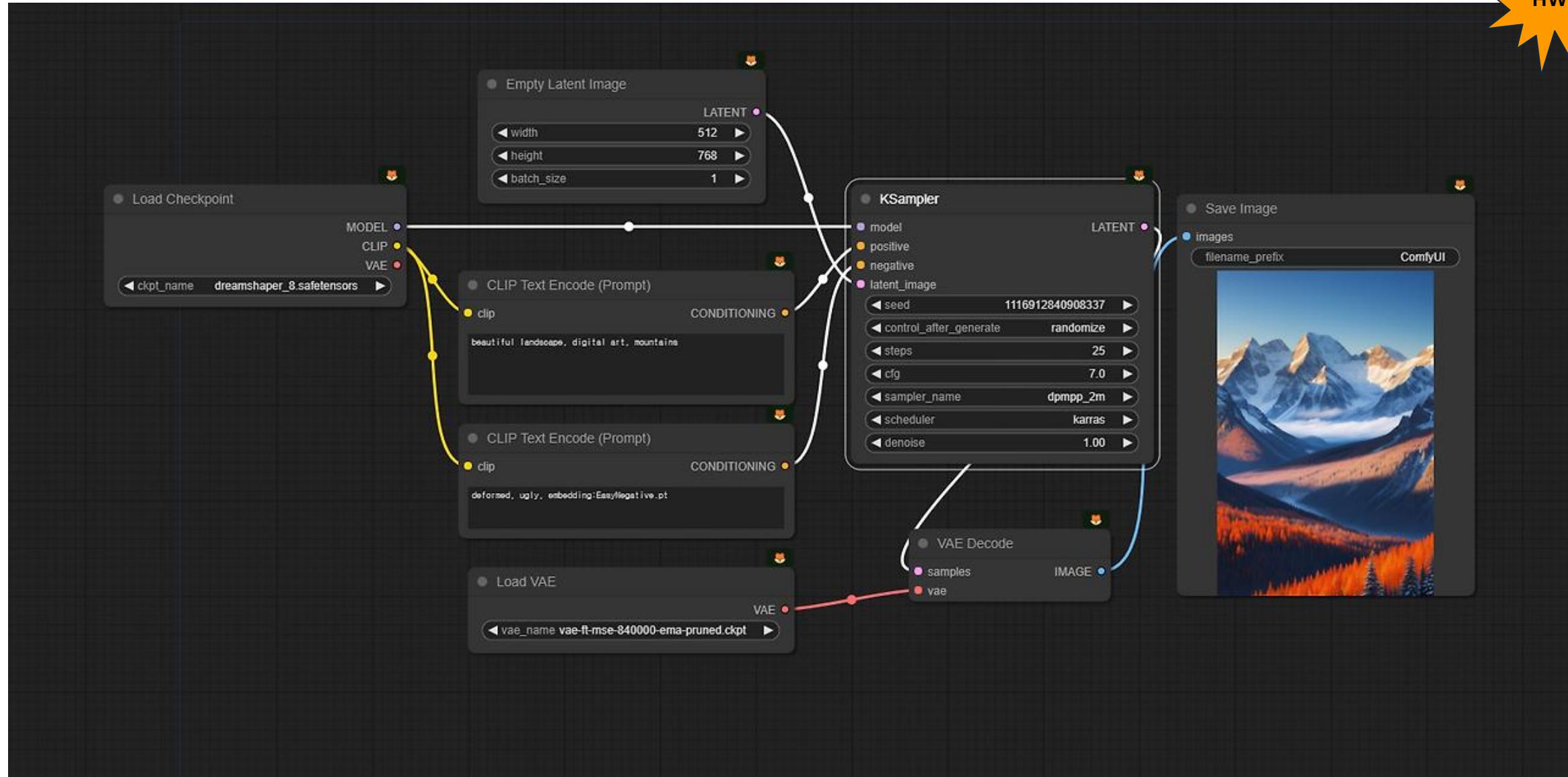
2. U-Net (Denoiser)

- 잠재 공간에서 노이즈를 점진적으로 제거하는 핵심 모델
- cross-attention을 통해 텍스트 조건(프롬프트) 반영

3. VAE (Decoder)

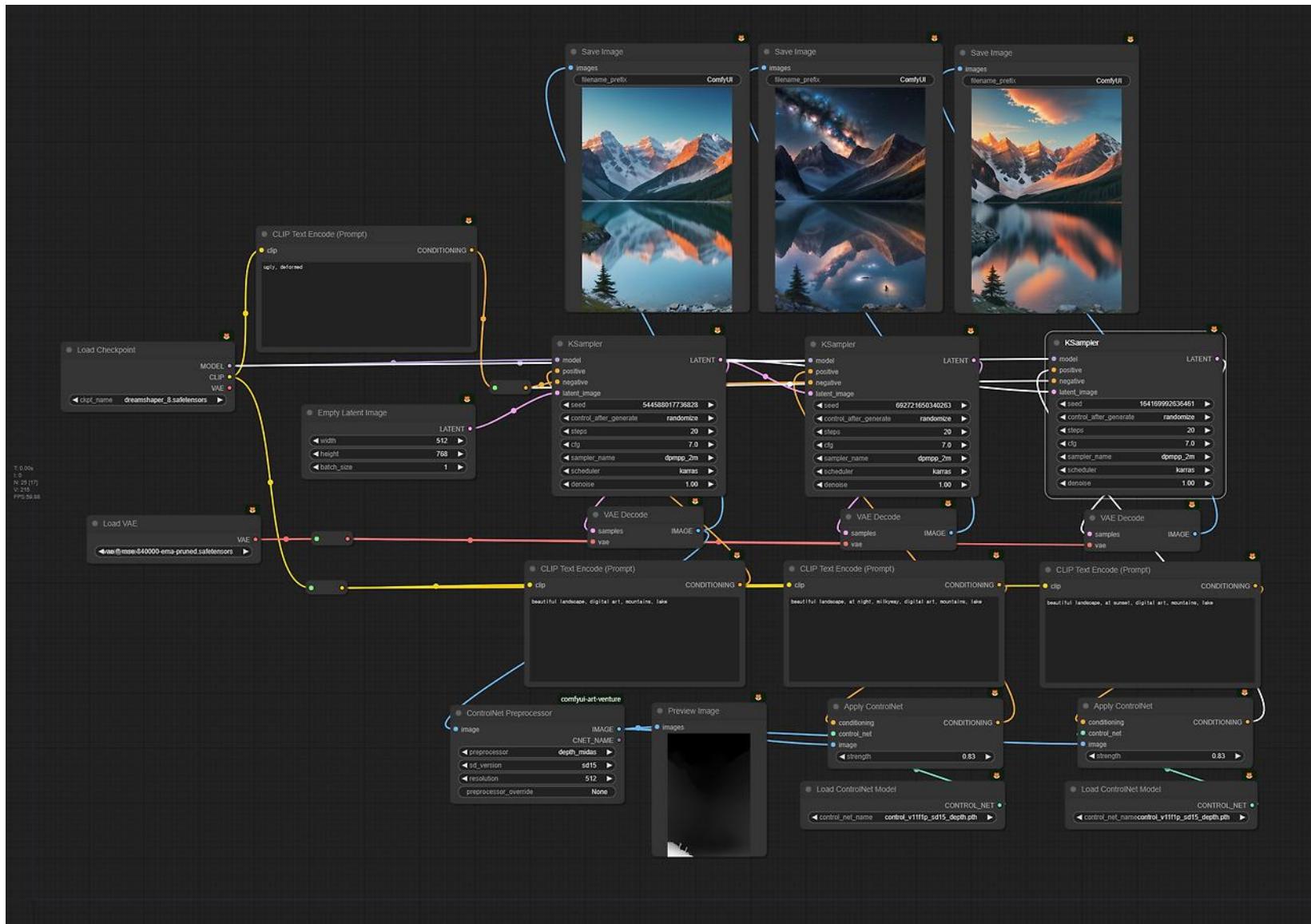
- 마지막에 잠재 표현을 다시 픽셀 공간으로 디코딩
- 고해상도 이미지를 출력





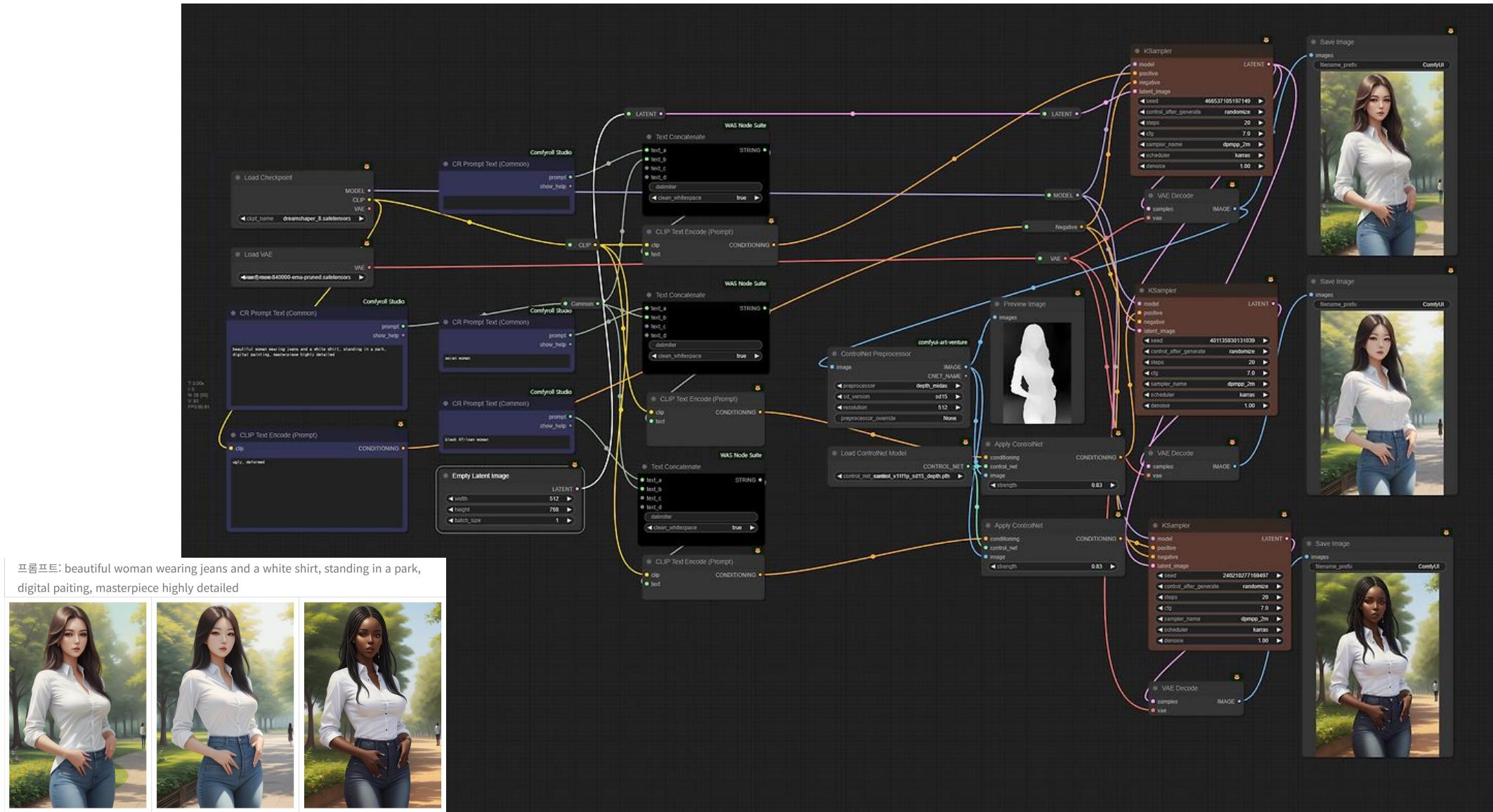
아래는 아름다운 산이 있는 풍경을 생성하는 가장 기본적 워크플로입니다.

위에 있는 워크플로에서 맨 오른쪽의 [Preview Image] 노드를 [Save Image]노드로 바꿨고, 아래쪽에 [VAE Decode]노드에서 별도의 VAE를 가져오도록 [Load VAE] 노드를 추가하였습니다. 참고로 이 워크플로는 그냥 다운로드 받아서 ComfyUI 화면에 Drag&Drop하면 사용할 수 있습니다.



이 워크플로는 첫번째 사진을 생성한 후, 이 사진을 참조하여 콘트롤넷 Depth 맵을 만들고, 두번째 세번째 이미지에 적용하는 방법입니다.
해당 내용은 아래와 같습니다. 먼저 처음 생성된 이미지에서 이미지를 가져다가 [ControlNet Preprocessor] 노드에서 "depth_midas" 전처리기를 수행해 Depth map을 생성합니다([Preview Image] 노드에 이 Depth map이 나타납니다).

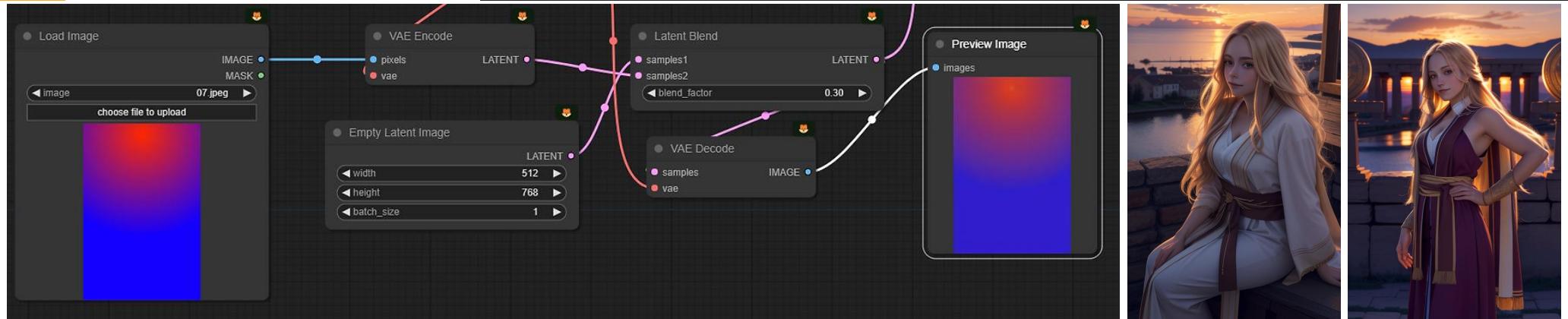
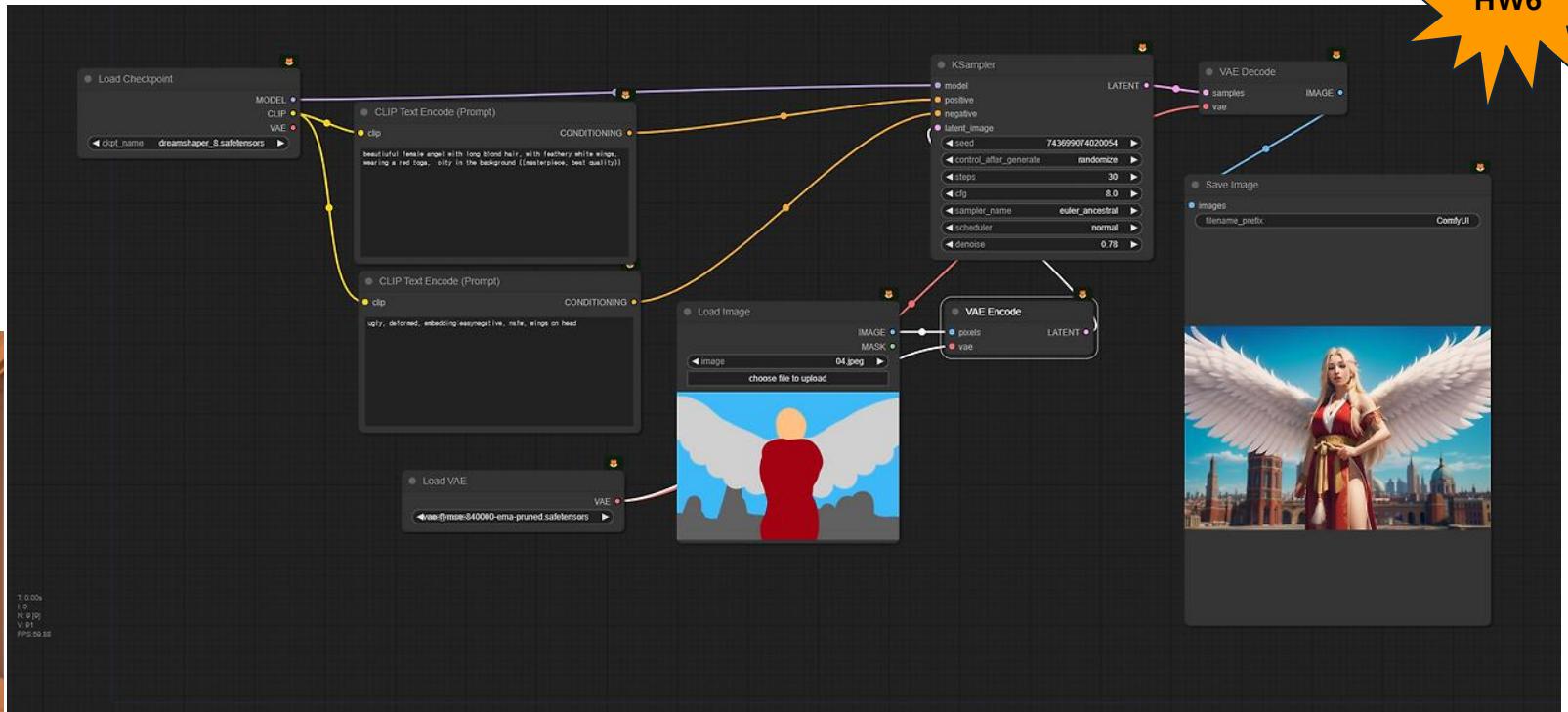
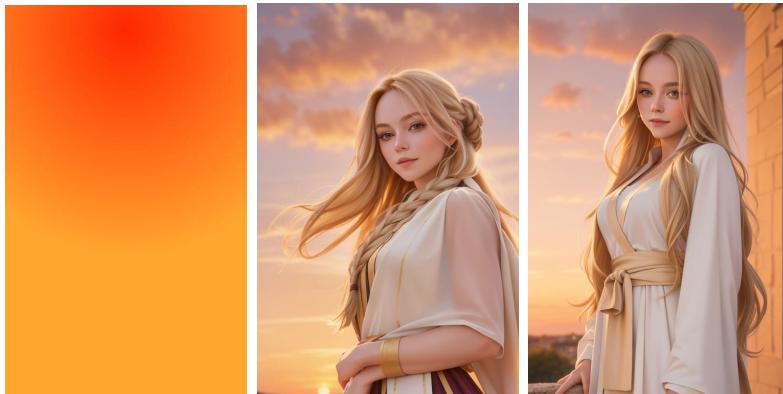
HW6



프롬프트를 공통 프롬프트와 개별적으로 적용하는 프롬프트를 별도로 분리해서 입력하고 이를 연결(Concatenate)한 워크플로입니다. 프롬프트는 아래와 같습니다. 첫번째는 beautiful woman, 두 번째는 beautiful asian woman, 마지막은 beautiful black African woman 으로 바꿨습니다. 보시는 것처럼 거의 동일한 배경, 거의 동일한 자세 및 복장으로 인종만 다르게 생성되었습니다.

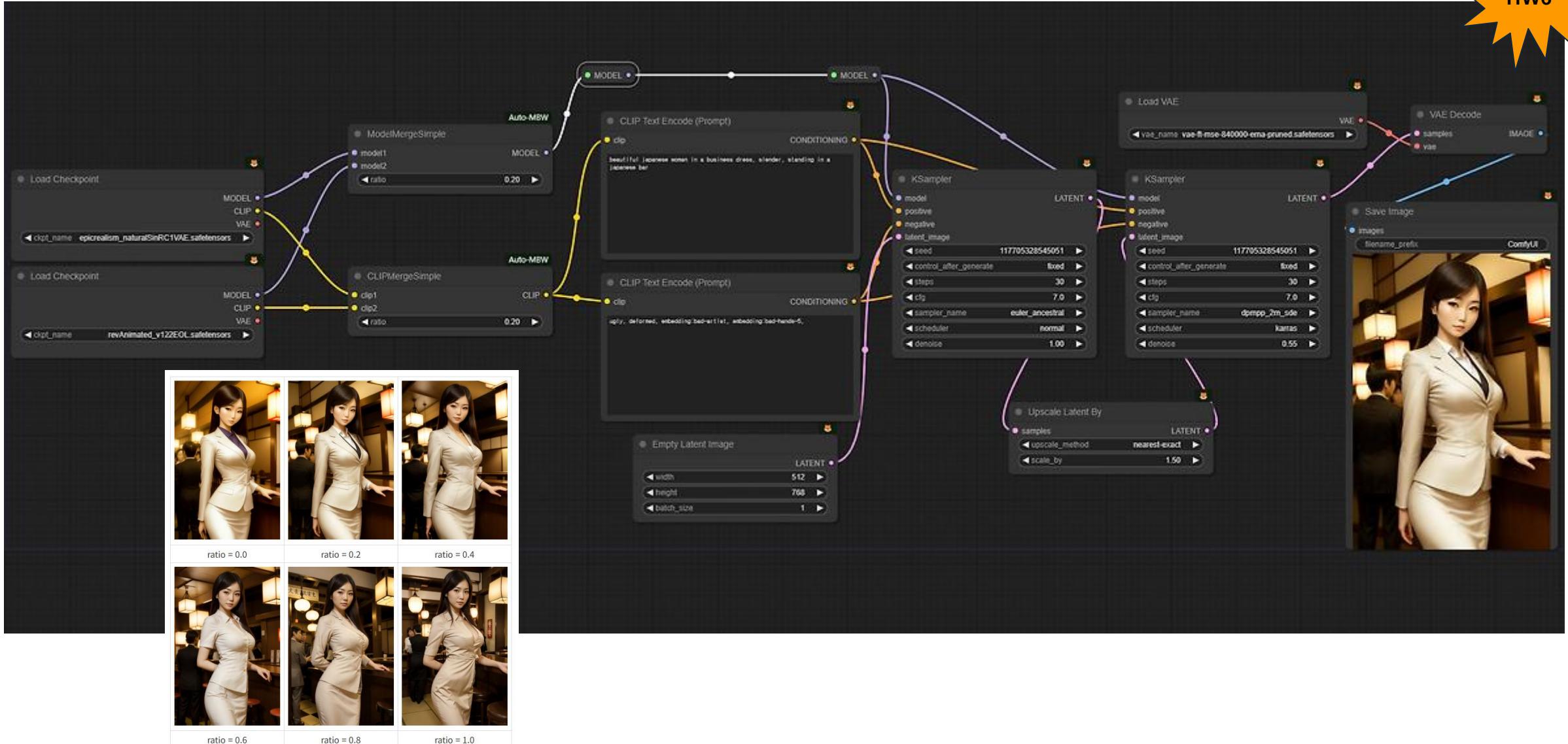
HW6

전체적으로 위쪽은 적황색, 아래쪽은 파란색이 지배적이기는 하지만, 색감이 훨씬 풍부해졌음을 알 수 있습니다. 그래디언트 이미지를 바꿔주기만 하면 다양한 이미지를 얻을 수 있습니다. 아래의 그래디언트 이미지를 사용하면 아래와 같은 이미지가 생성됩니다.

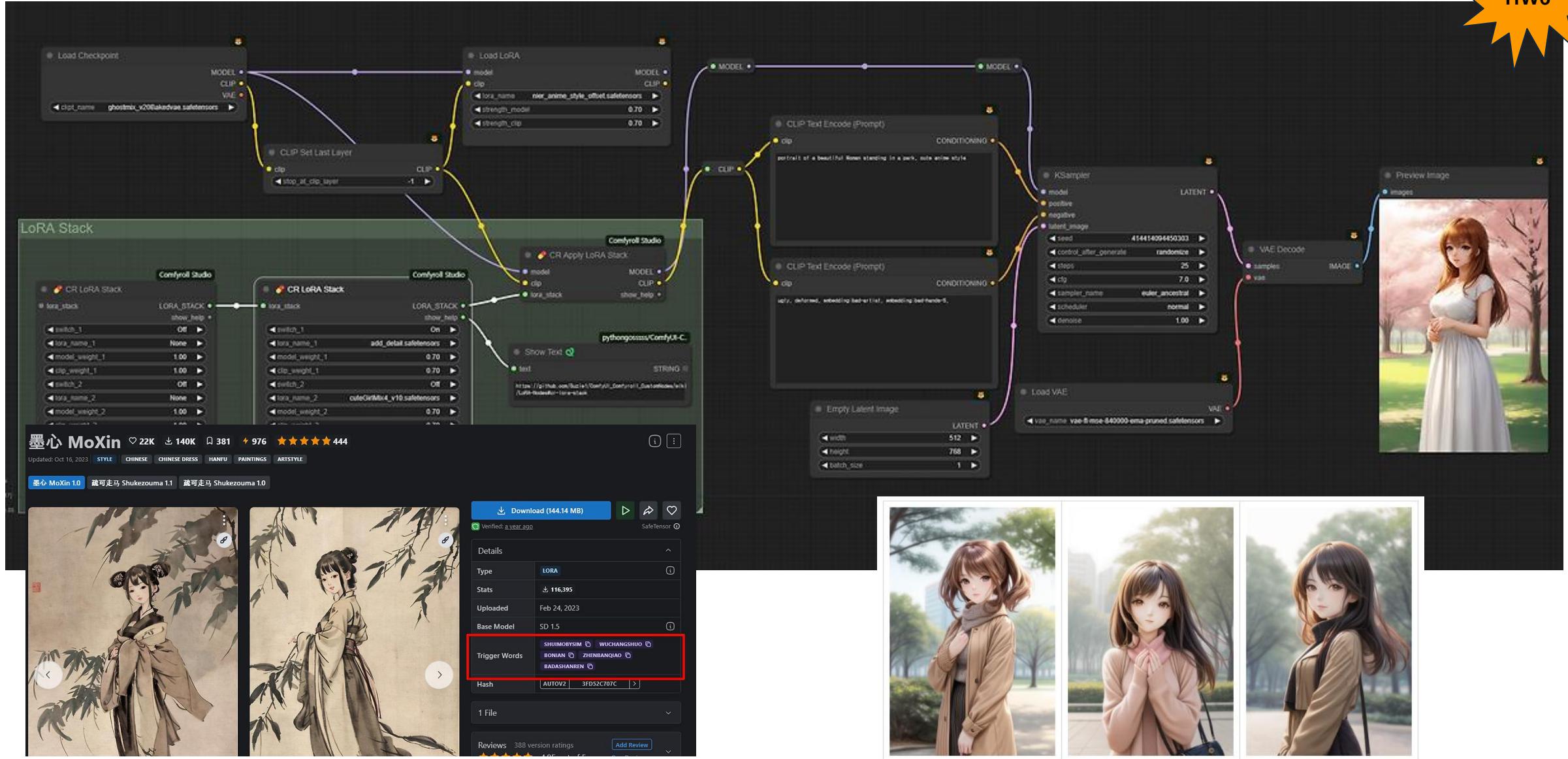


img2img 워크플로가 txt2img 워크플로와 다른 점은, txt2img에서는 [Empty Latent Image] 노드로부터 그냥 무작위 이미지를 소스로 사용하지만, img2img에서는 [Load Image]로부터 이미지를 입력받은 후, 이를 [VAE Encode] 노드에서 잠재이미지를 변환해 이를 소스로 사용한다는 점만 다릅니다. 이를 통해, txt2img에서는 어떤 이미지가 생성될지 전혀 모릅니다. img2img에서는 전체적인 구도와 색을 지정할 수 있다는 장점이 있습니다.

76

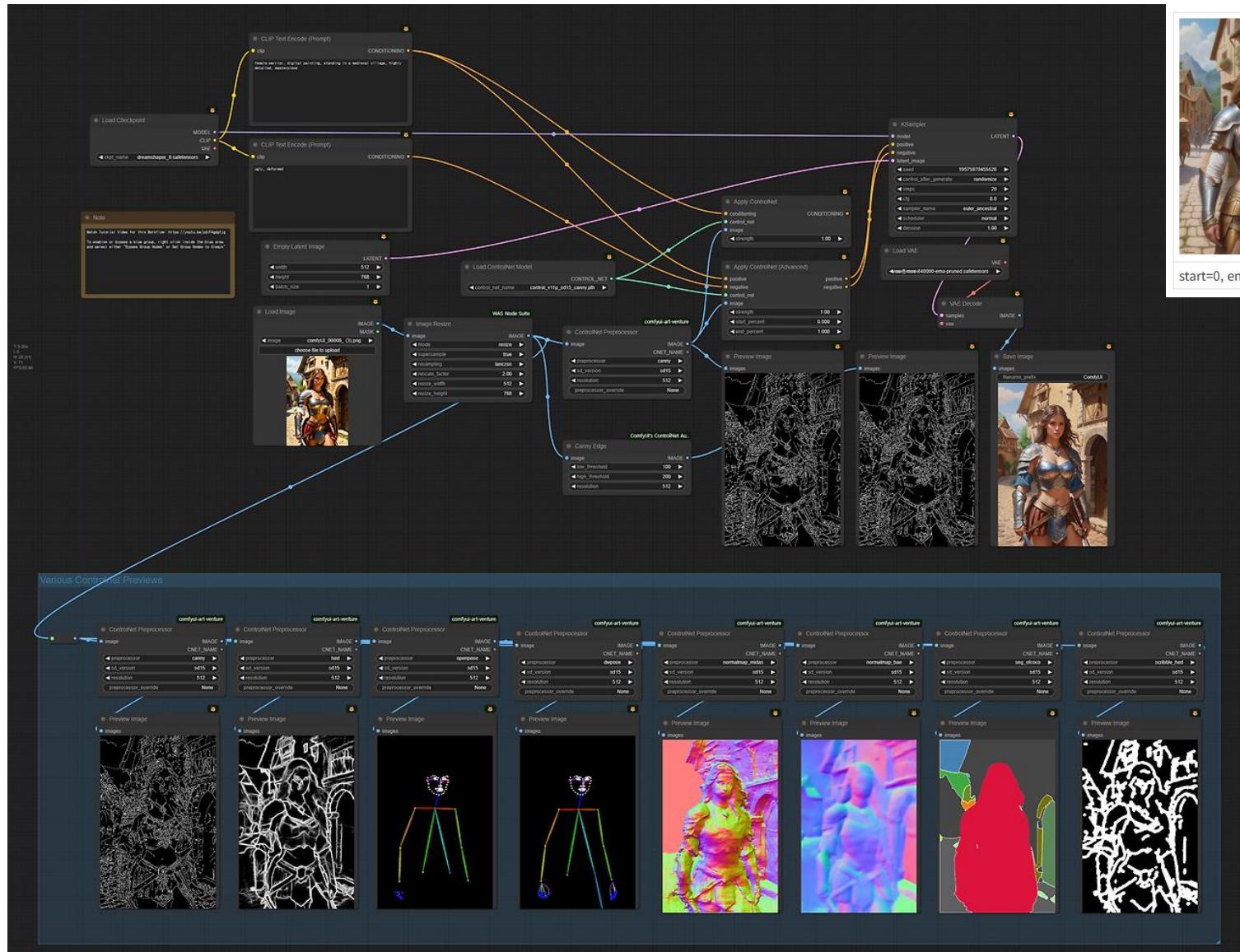


이 워크플로에서도 [KSampler]를 두번 실행시킵니다. 아래에서 좌우를 비교해보면, 샘플러를 다른 것(euler_ancestral과 dpmpp_2m_sed)을 사용하였고, 또한 잡음제거 강도가 처음은 1.0, 나중은 0.55를 적용하였음을 알 수 있습니다. 아래는 ratio를 달리하면서 두개의 모델을 합쳐서 생성한 결과입니다. 프롬프트는 "beautiful japanese woman in a business dress, slender, standing in a japanese bar" 입니다. 여기에서 ratio는 model1으로 지정된 모델이 차지하는 비중으로서, 0.2로 지정하면 epicRealism : revAnimate = 2 : 8 이라는 ⁷⁷ 뜻입니다. 아래에서 두 모델이 차지하는 비율이 달라지는 효과를 보실 수 있을 것입니다.



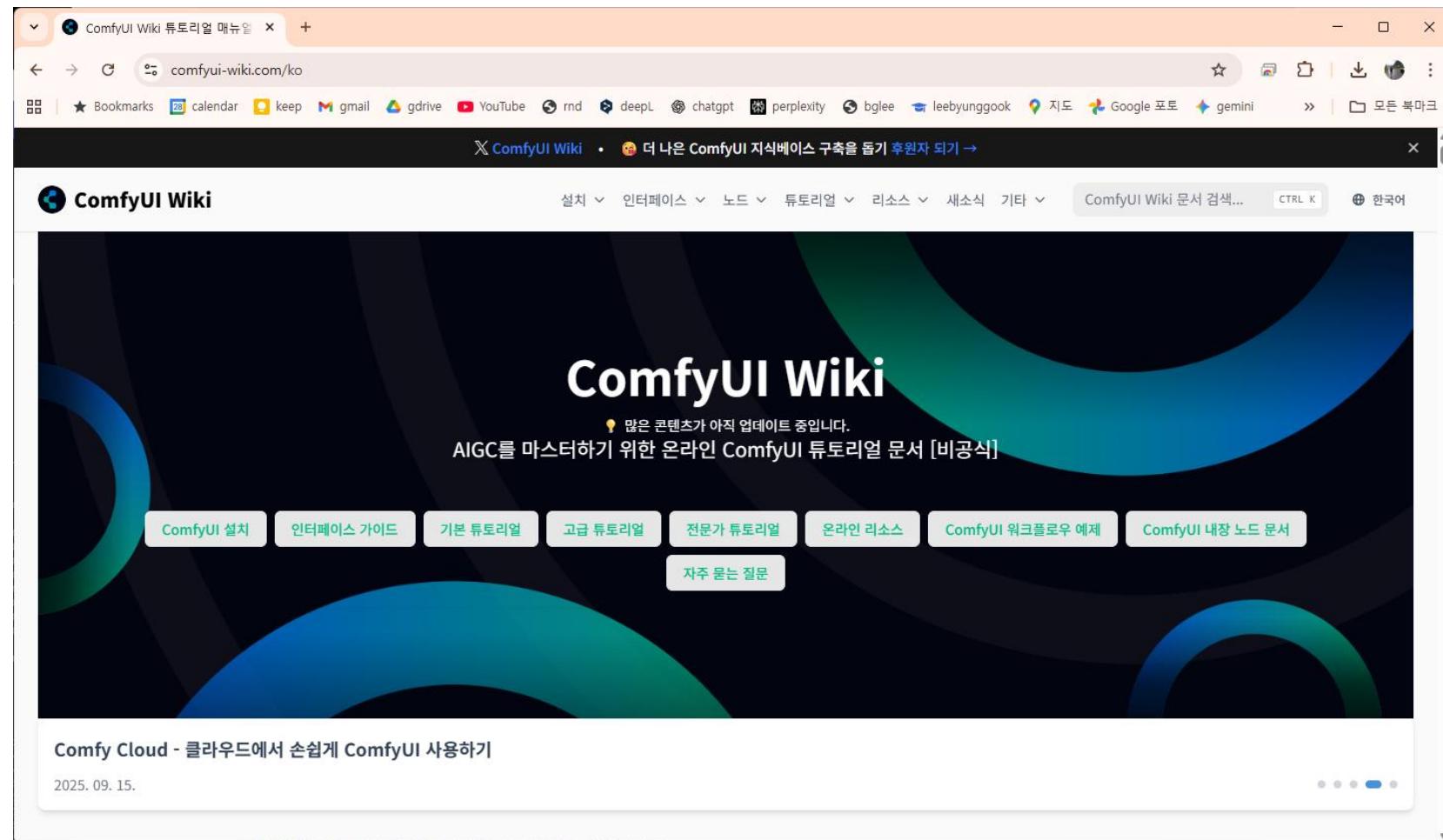
LoRA는 LoRA에 따라 성능이 발휘하는 조건이 다릅니다. 가장 중요한 것은 체크포인트 모델과의 조화입니다. 해당 LoRA의 스타일을 잘 드러내주는 체크포인트 모델이 있다는 것입니다. 또한 아래 그림에서 보는 것처럼 LoRA 파일을 발현시켜주는 "Trigger Words"가 따로 있는 경우가 있는데, 이때는 프롬프트에 해당 단어가 포함되어야 합니다. 아울러, CFG 값이나 가중치(weight) 등도 중요할 수 있습니다. 이러한 사항은 대개 LoRA 파일 사이트에 적혀 있으니 가능하면 찾아서 적용하는 것이 좋습니다.

HW6

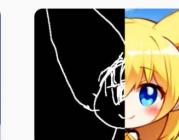
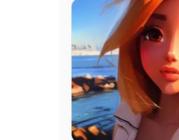


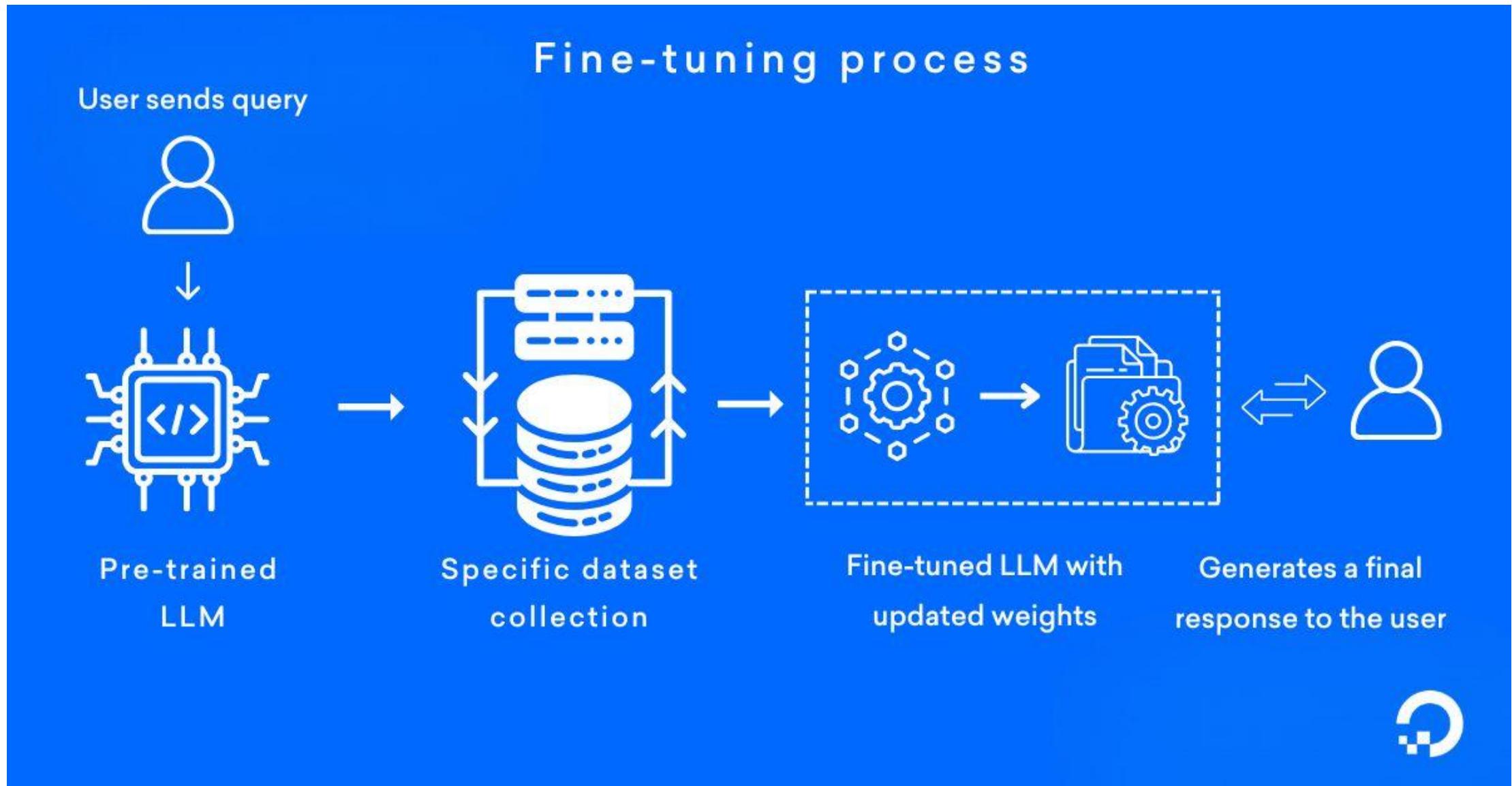
ControlNet은 참조 이미지의 자세와 구도를 복사하여 새로운 이미지를 만드는, Stable Diffusion에서 매우 중요한 모델입니다. 즉, ControlNet을 이용하면 참조 이미지와 비슷하지만, 다른 스타일의 이미지를 생성할 수 있습니다. ComfyUI에서 콘트롤넷을 사용하는 방법은 [여기](#)를 참고하시기 바랍니다. 아래는 위의 워크플로를 사용해 생성한 이미지입니다. 왼쪽은 원본 이미지이고, 나머지 두개는 생성한 이미지입니다. 이 경우에는 외곽선을 추출하는 Canny 알고리즘을 사용했기 때문에 구도가 거의 동일하고, 색감이나 스타일만 약간 다릅니다.

ComfyUI Wiki



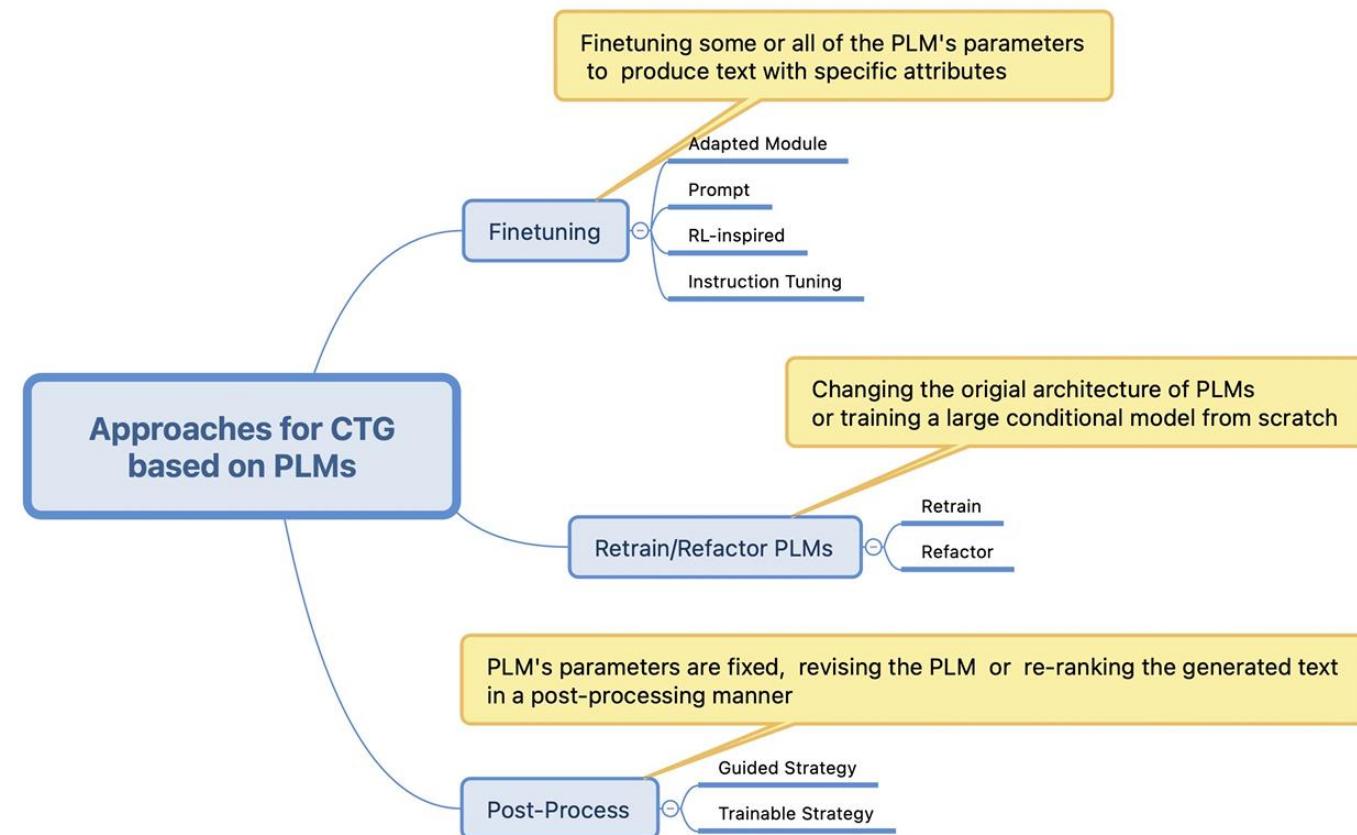
ComfyUI Workflow

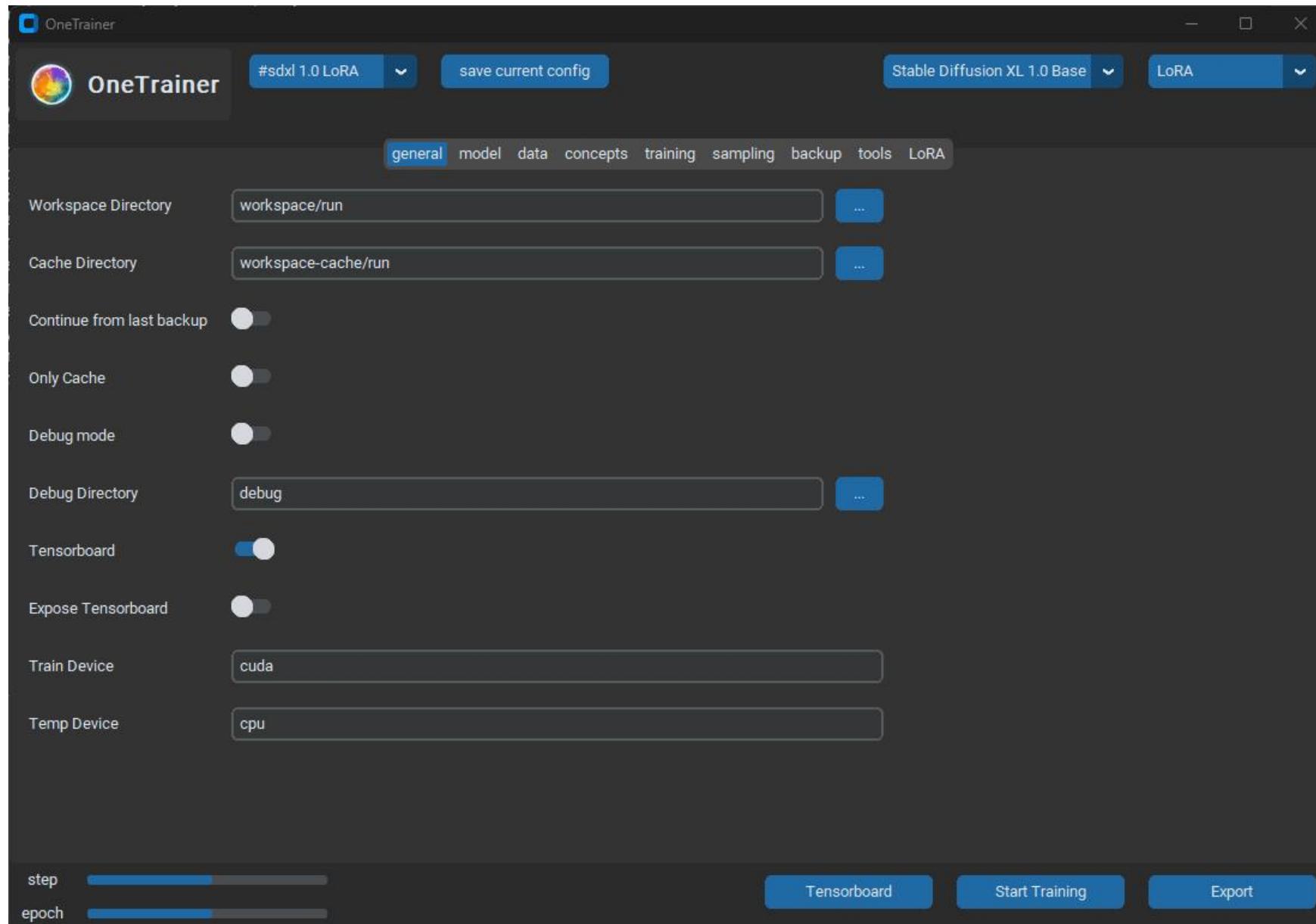
| | | | | | | | |
|---|--|--|---|---|--|--|---|
|  1. Img2Img Img2Img |  2. 2 Pass Txt2Img Text2Img Upscale |  3. Inpaint Inpaint |  4. Area Composition 이 문서는 ComfyUI의 영 역 구성 예제를 소개합… Cascade |  13. Stable Cascade UnCLIP |  14. UnCLIP UnCLIP |  15. Hypernetworks Hypernetwork |  16. Glichen |
|  5. Upscale Models 이 문서는 ComfyUI의 업 스케일 모델 예제를 소… 업스케일 |  6. LoRA Lora |  7. ControlNet ControlNet |  8. Noisy Latent Composition 3d |  17. 3D Examples 이 문서는 ComfyUI에서 3D 생성의 몇 가지 예제… Video |  18. Video 이 문서는 ComfyUI의 이 미지에서 비디오로의 예… Video |  LCM 예제 (LCM Examples) Lcm |  ComfyUI SDXL Turbo 예제 Turbo |
| | | | |  9. Textual Inversion… Embeddings |  10. Edit Models |  11. Model Merging Model merging |  12. SDXL Sdxl |



LoRA: Low-Rank Adaptation of Large Language Models

LoRA의 간단하지만 핵심을 짜르는 아이디어로 인해 현존하는 거대 모델들을 적은 비용으로 파인튜닝할 수 있게 되었습니다. 그리고 이러한 LoRA를 개선하기 위해 많은 수학적, 경험적 아이디어들이 쏟아져 나오고 있습니다. 인공지능을 더욱 더 잘 활용하기 위해서는 필수적으로 pre-trained 모델을 목적에 맞게 조금씩 파인튜닝해야 할 것이고 따라서 이러한 PEFT 기법은 모든 AI 엔지니어나 연구원들이 반드시 알아야 하는 필수적인 기법이 될 것입니다. LoRA, LoHa, LoKr은 그 중 가장 대표적인 방법들이고 어떤 원리로 파라미터 효율적인 파인튜닝을 가능하게 했는지 알아보았습니다.





OneTrainer is a one-stop solution for all your stable diffusion training needs.

<https://medium.com/@alivesin/lora-training-step-3-training-your-first-lora-da8e54a33df2>

<https://huggingface.co/collections/linoyts/datasets-for-lora-training-66aba176e77d184cb91cfa65>

https://www.reddit.com/r/StableDiffusion/comments/1gvp073/a_personal_experience_guide_to_training_sdxl/

<https://www.youtube.com/watch?v=a03BNvDVRdQ>

<https://www.patreon.com/posts/onetrainer-fine-101606471>

<https://www.youtube.com/watch?v=0t5l6CP9eBg>

1. [Python 3.10.x, Git, CUDA 12.x](#)

2. Add python to system path

3. make work directory SdFineTuning

4. [Download datasets for train Lora yarn_art_style.zip](#)

5. **Install Onetrainer - Windows Powershell**

a. cd work directory

b. git clone <https://github.com/Nerogar/OneTrainer.git>

c. cd Onetrainer

d. python -m venv venv

e. Set-ExecutionPolicy RemoteSigned -Scope CurrentUser

f. venv\Scripts\activate

g. pip install -r requirements.txt

h. .\update.bat

i. .\start-ui.bat

6. **Install ComfyUI - Windows Powershell**

a. cd work directory

b. git clone <https://github.com/comfyanonymous/ComfyUI.git>

c. cd ComfyUI

d. python -m venv venv

e. Set-ExecutionPolicy RemoteSigned -Scope CurrentUser

f. venv\Scripts\activate

g. pip install torch torchvision torchaudio --index-url <https://download.pytorch.org/whl/cu121>

h. pip install -r requirements.txt

7. [Download checkpoint](#)

a. [Sd_xl_base_1.0.safetensors](#)

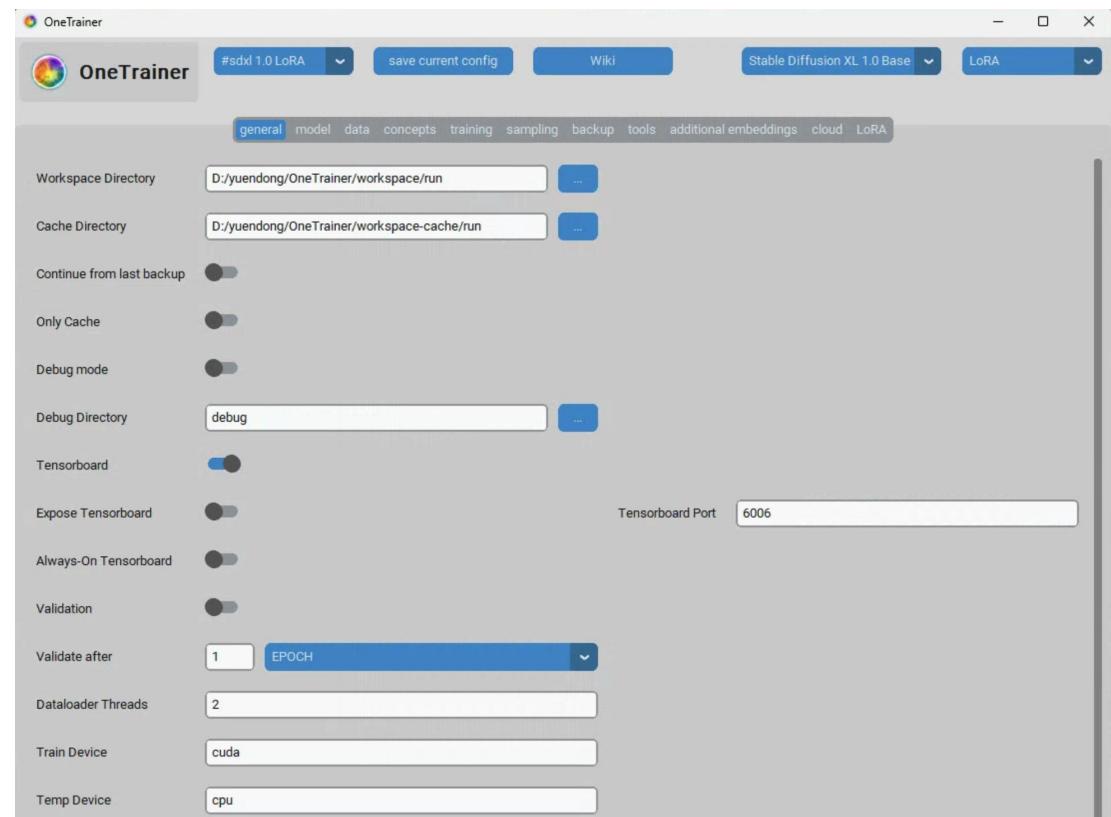
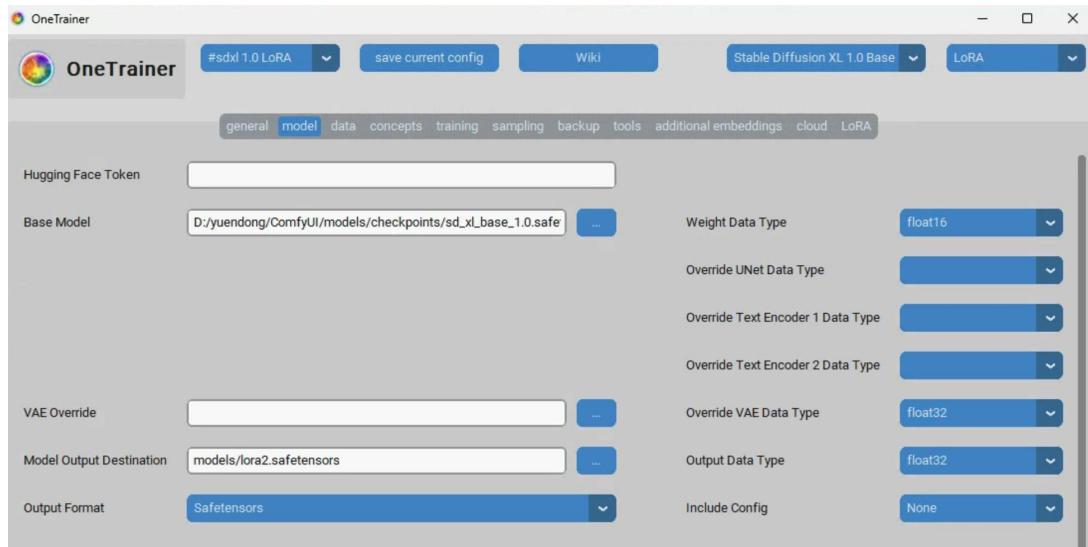
b. Move to ComfyUI/models/checkpoints/sd_xl_base_1.0.safetensors



Onetrainer

OneTrainer/start-ui.bat

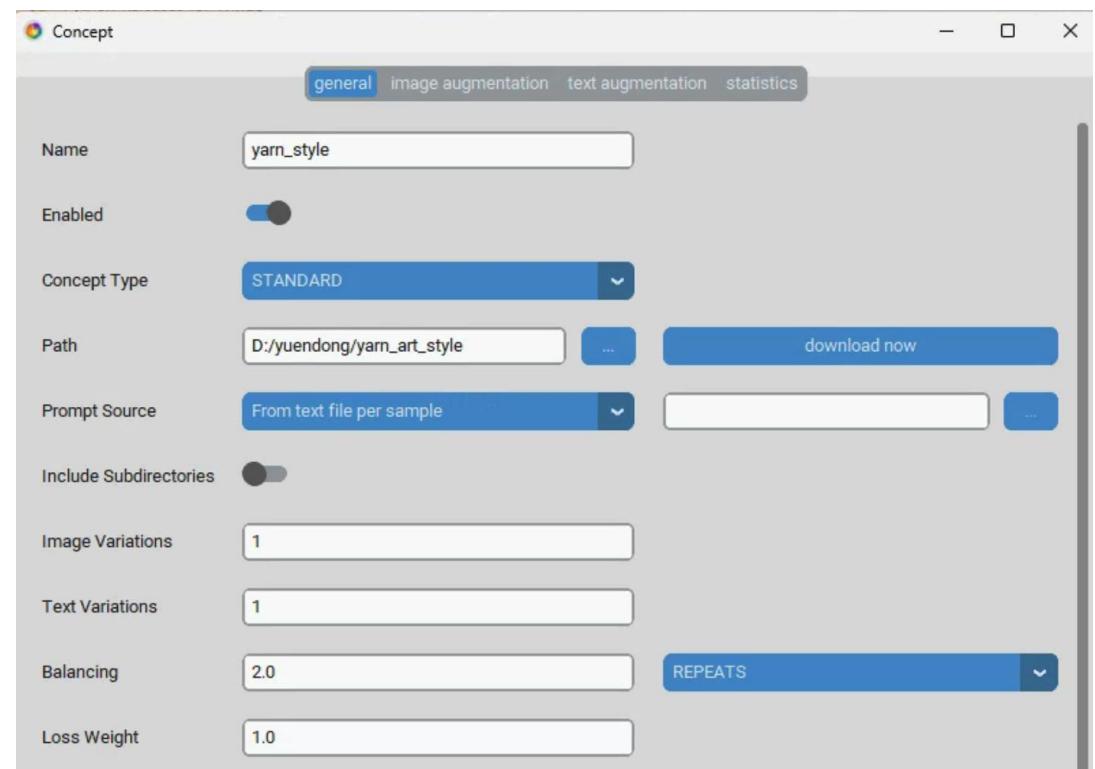
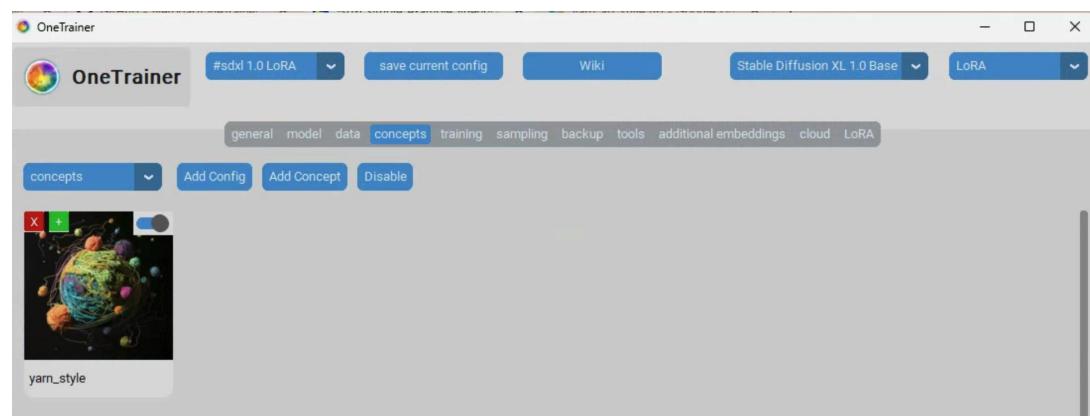
1. Lora mode, [#sdxl 1.0 LoRA], [LoRA]
2. General Tab
 - a. Create folder
 - i. OneTrainer/workspace/run
 - ii. OneTrainer/workspace-cache/run
 - b. Set Workspace Directory
3. Model tab, load Base Model



Onetrainer

4. Concepts Tab, Add Concept, Click concept to edit

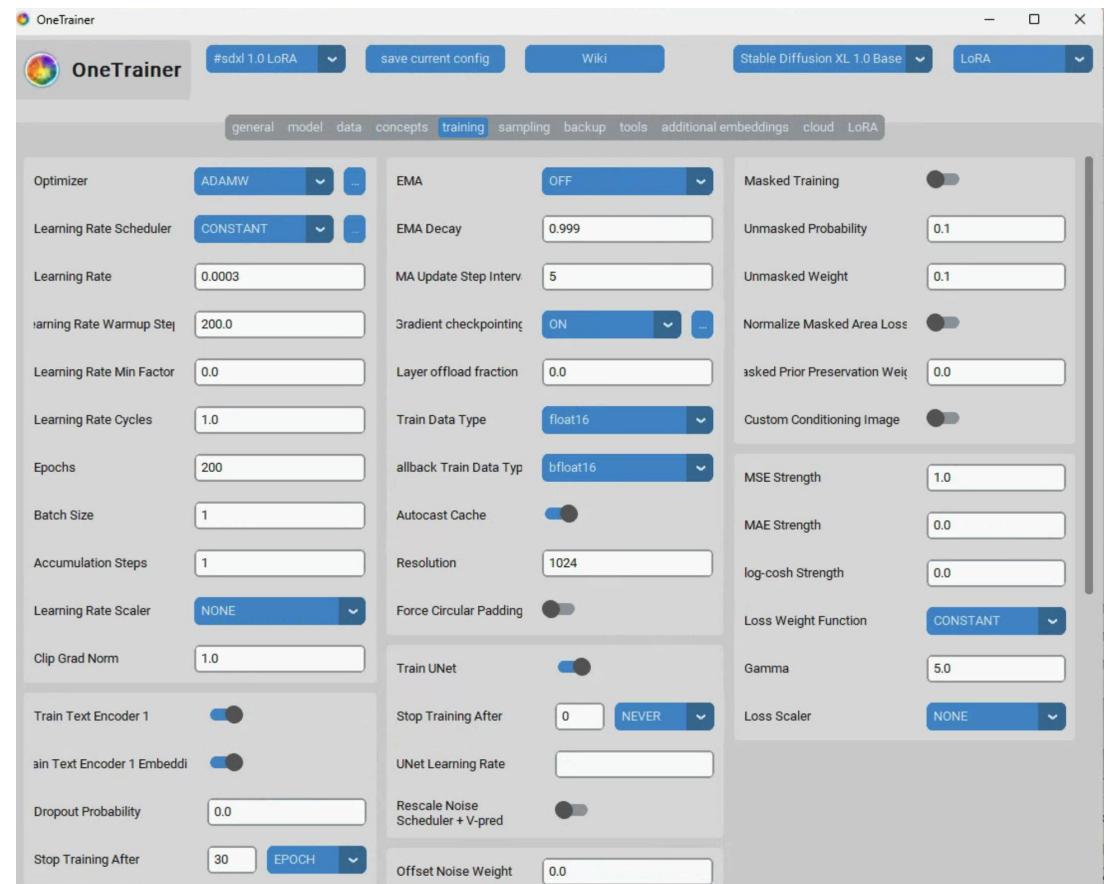
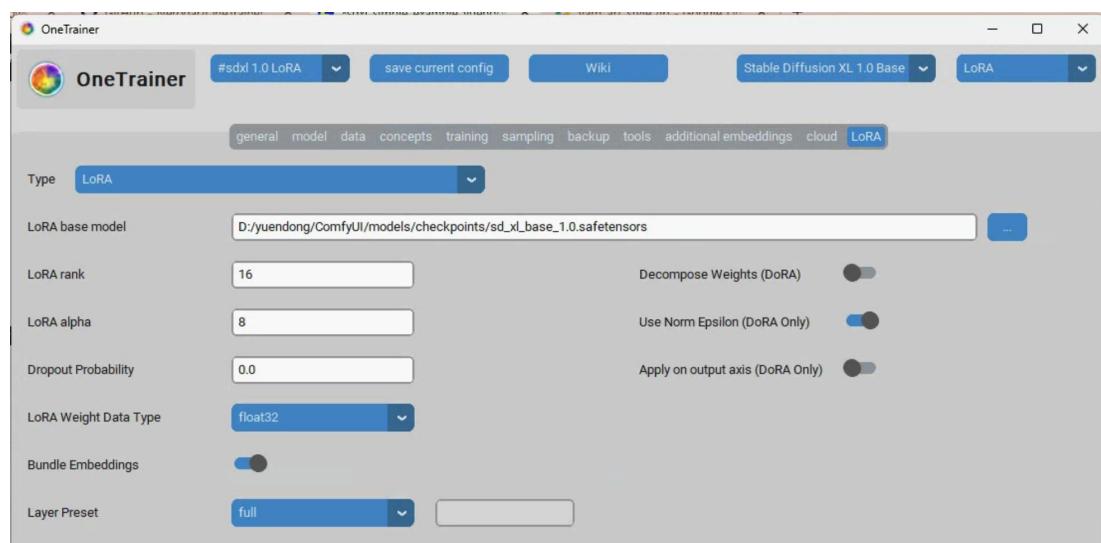
- a. Input concept name
- b. Choose datasets path
- c. Set Balancing



Onetrainer

5. Training

- a. Epochs - 200
- b. Batch Size - 1 : Use 8G VRAM
- c. Enable Train Text Encoder 1
- d. Enable Train UNet



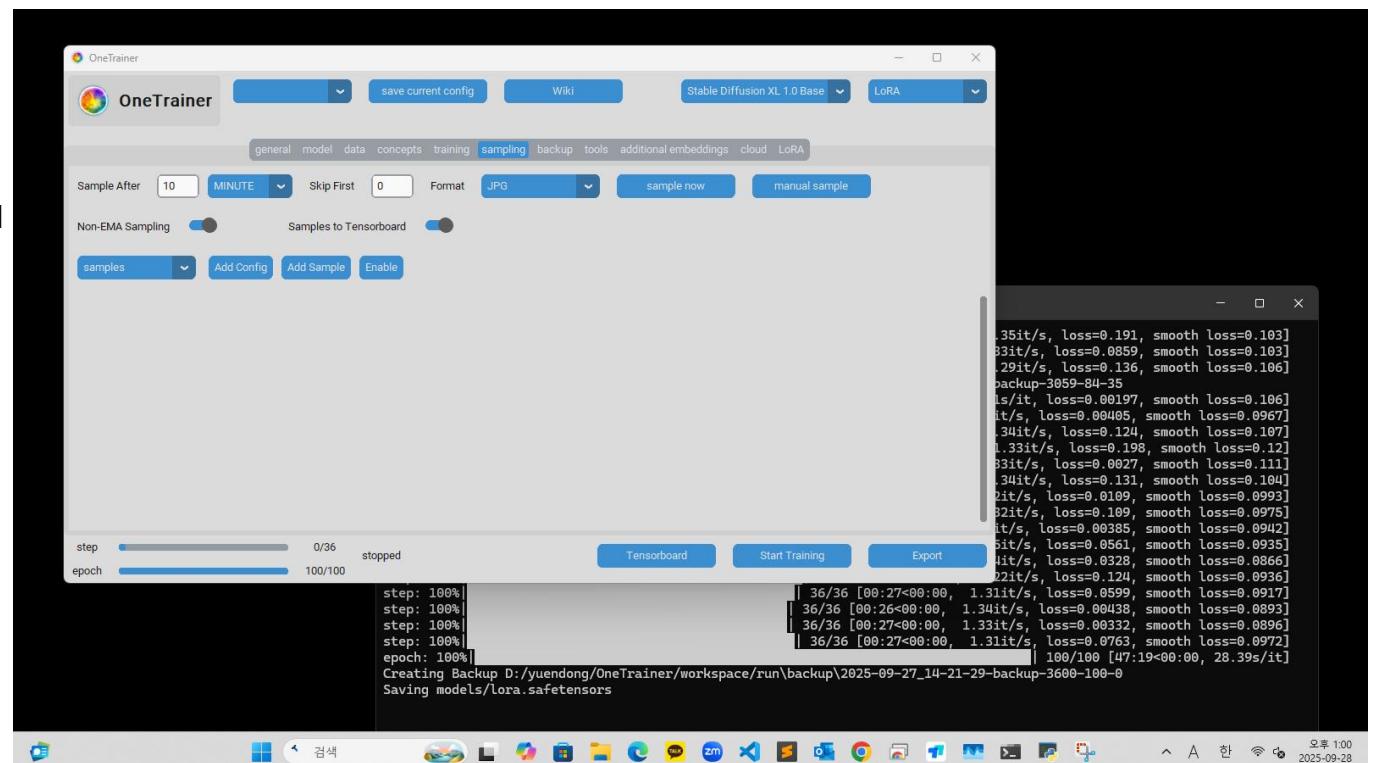
Onetrainer

Start Training

- Weight Data Type: float16
- Override VAE Data Type: float16
- Output Data Type: float16
- Epochs: 50-200
- Batch Size: 1-4,
- EMA: Enabled
- EMA Update Step Interval: 1~5
- Train Text Encoder 1: Option, will train text prompt when enabled
- Train UNet: Must enabled if train lora
- LoRA Weight Data Type: float32

Test Lora with Comfyui

- Copy Lora file
 - Move to ComfyUI/models/loras/my_style_lora.safetensors
- Load SDXL Workflow



SeaArt – [SPOILER] Gear Fifth Luffy LoRA

Stable Diffusion

- 개별 LoRA 모듈로 캐릭터 스타일, 그림체, 특정 인물 재현 등을 학습.
- 크리에이터들이 공유·교환하며, 모델 전체가 아닌 소형 LoRA 모듈(수 MB)로 배포 가능.
- “LoRA 합성” 기능을 활용해 여러 스타일을 동시에 적용하기도 함.

[SPOILER] Gear Fifth Luffy (One Piece) LoRA

Making models can be expensive. Do you like what I do? Consider supporting me on Patreon  or feel free to buy me a coffee 

[SPOILER FOR ONE PIECE CHAPTER 1044]

Concept for Gear 5th. Should apply well to Luffy, but also to any other character. This is not too overfitted on the trigger gear fifth, meaning you can change hair color and clothes if you want. You can also have different characters use Gear 5. Lower the weight if you want to get less of a "drawing" effect.

The dataset, however, was quite small, so the model picked up a bit of "cartoon style". However, if you know the origin of this power, this is not a problem. Not gonna explain more to avoid spoilers, but it's like when you say "it's not a bug, it's a feature".

Tested on AnyLoRA and NED. Also using my Wano style LoRA.

Enjoy.

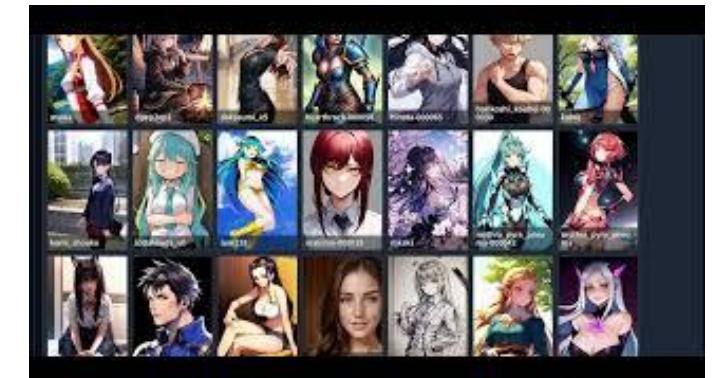
How to use LoRA's in auto1111:

Update webui (use git pull like here or redownload it)

Copy the file to stable-diffusion-webui/models/lora

Select your LoRA like in this video

Make sure to change the weight according to the instructions (by default it's :1)



SDXL LoRA DreamBooth - chinh102/onepiece-fina...

SDXL LoRA DreamBooth - chinh102/onepiece-fina...



Prompt [🔗](#)
Anime illustration of Luffy Gear 5: white cloud-like hair, open white jacket, purple shorts. Dynamic pose against night sky with full moon or clouds. Joyful...



Prompt [🔗](#)
Anime illustration of Luffy Gear 5: white cloud-like hair, open white jacket, purple shorts. Dynamic pose against night sky with full moon or clouds. Joyful...



Prompt [🔗](#)
Anime illustration of Luffy Gear 5: white cloud-like hair, open white jacket, purple shorts. Dynamic pose against night sky with full moon or clouds. Joyful...

Model description

These are chinh102/onepiece-fina... LoRA adaption weights for stabilityai/stable-diffusion-xl-base-1.0.

Download model

Use it with UIs such as AUTOMATIC1111, Comfy UI, SD.Next, Invoke

LoRA: download [on... here](#)

- Place it on your models/Lora folder.
- On AUTOMATIC1111, load the LoRA by adding <lora:on...> to your prompt. On ComfyUI just [load it as a regular LoRA](#).

Anime illustration of Luffy Gear 5: white cloud-like hair, open white jacket, purple shorts. Dynamic pose against night sky with full moon or clouds. Joyful expression, wide grin. Cartoonishly flexible body, stretched limbs. Glowing effects around figure. Energetic movement, mid-jump or playful stance. Contrast between illuminated character and dark background.

- Embeddings: download [on... here](#).
 - Place it on your embeddings folder
 - Use it by adding onepiece-fina... emb to your prompt. For example, a photo of onepiece-fina... emb luffy gear 5 (you need both the LoRA and the embeddings as they were trained together for this LoRA)

Use it with the [diffusers library](#)

```
from diffusers import AutoPipelineForText2Image
import torch
from huggingface_hub import hf_hub_download
from safetensors.torch import load_file
```

```
pipeline = AutoPipelineForText2Image.from_pretrained('stabilityai/stable-diffusion-xl-base-1.0')
pipeline.load_lora_weights('chinh102/onepiece-fina...', weight_name='pytorch_lo...')
embedding_path = hf_hub_download(repo_id='chinh102/onepiece-fina...', filename='state_dict = load_file(embedding_path)
pipeline.load_textual_inversion(state_dict["clip_1"], token=[<s0>, <s1>], tex...
pipeline.load_textual_inversion(state_dict["clip_g"], token=[<s0>, <s1>], tex...
```

```
image = pipeline('Anime illustration of Luffy Gear 5: white cloud-like hair, open
```

For more details, including weighting, merging and fusing LoRAs, check the [documentation on loading LoRAs in diffusers](#)



Trigger words

To trigger image generation of trained concept(or concepts) replace each concept identifier in your prompt with the new inserted tokens:

to trigger concept T0K → use <s0><s1> in your prompt

Details

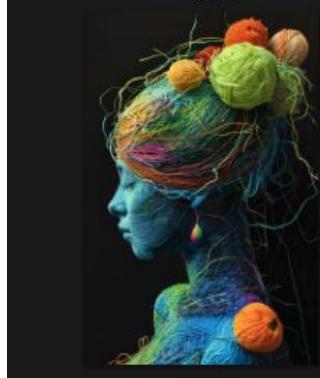
All [Files & versions](#)

The weights were trained using [diffusers Advanced Dreambooth Training Script](#).

LoRA for the text encoder was enabled. False.

Pivotal tuning was enabled: True.

Special VAE used for training: madebyollin/sdxl-vae-fp16-fix.



Prompt

Serene raven-haired woman,
moonlit lilies, swirling botanicals,
yarn art style

Prompt

a puppy in a pond, yarn art style

Prompt

Ornate fox with a collar of autumn
leaves and berries, amidst a
tapestry of forest foliage, yarn art
style



Prompt A green Cthulhu is rising from the blue sea in a great lightning storm, based on a story by H.P. Lovecraft, Paper Cutout Style

Prompt Kermit the frog, Paper Cutout Style

pill and candy mosaic style

Pill and Candy Mosaic Style [FLUX] 10 117 63 5 356

Updated: Jan 15, 2025 [STYLE](#) [MOSAIC](#) [CANDY](#) [PILL](#) [CANDIES](#)

V1

Have Pill and Candy Mosaic in your prompts

Norod78 pill-and-candy-mosaic-style-flux [use](#)

Test-to-Image Diffusers Norod78/PillAndCandyMosaic stable-diffusion lora template-lora migrated style mosaic candy pill candies [license](#): bespoke-lora-trained-license

Model card [Files and versions](#) [Community](#)

Pill and Candy Mosaic Style [FLUX]

Downloads last month: 27

Inference Providers: Pill and Candy Mosaic, cute cat

Pill and Candy Mosaic, cute cat

Pill and Candy Mosaic, Planet earth floating in space

Pill and Candy Mosaic, Planet earth floating in space

Pill and Candy Mosaic, A cute closeup on a Llama's head

Pill and Candy Mosaic, A cute closeup on a Llama's head

norod78 Joined Dec 8, 2022

453 5.2K

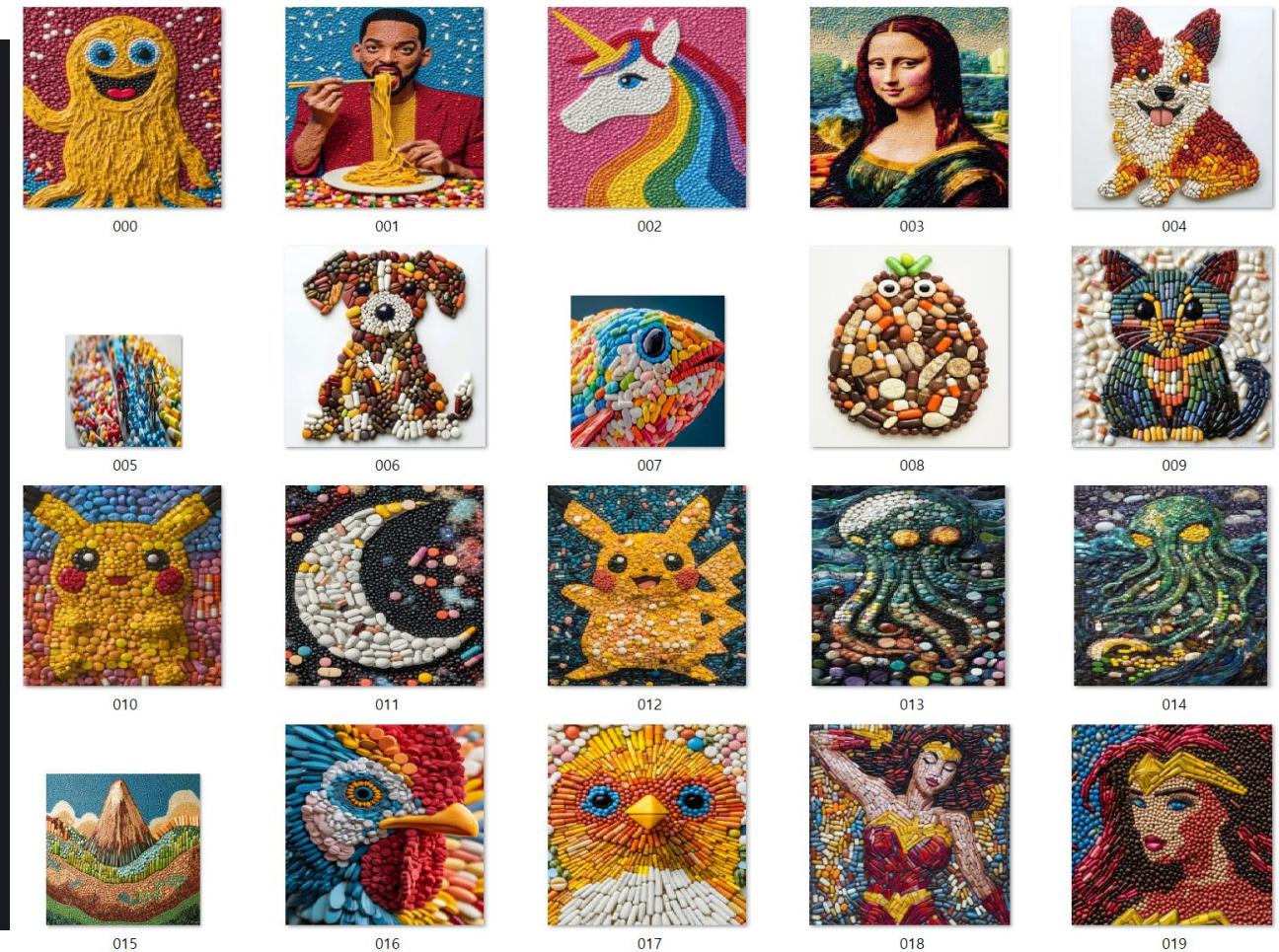
Dataset used to train Norod78/pill-and-candy-mosaic-style-flux

Base model: black-facet-labs/FLUX_1-dev

Adapter: Adapter [35904]

This model

CivitAI



Human + Meow

HuMeow LoRA 89 712 1.8K 14 3K

Updated: Jan 16, 2024 [CHARACTER](#) [CATS](#) [CAT](#) [FURRY](#) [HUMANOID CAT](#)

[SDXL-HuMeow-LoRA-r8](#) [SD15-HuMeow-LoRA](#)

As you all know, there are not enough cat images on the internet.

This is HuMeow (Human + Meow): An SDXL and SDv1.5 LoRAs which were designed to solve this problem by generating human(oid) cats as well as getting random cats to appear in scenes where the trigger word 'HuMeow' is used.

[Create](#) [Download](#) [Share](#) [Upvote](#) [Report](#)

Verified: 2 years ago SafeTensor ⓘ

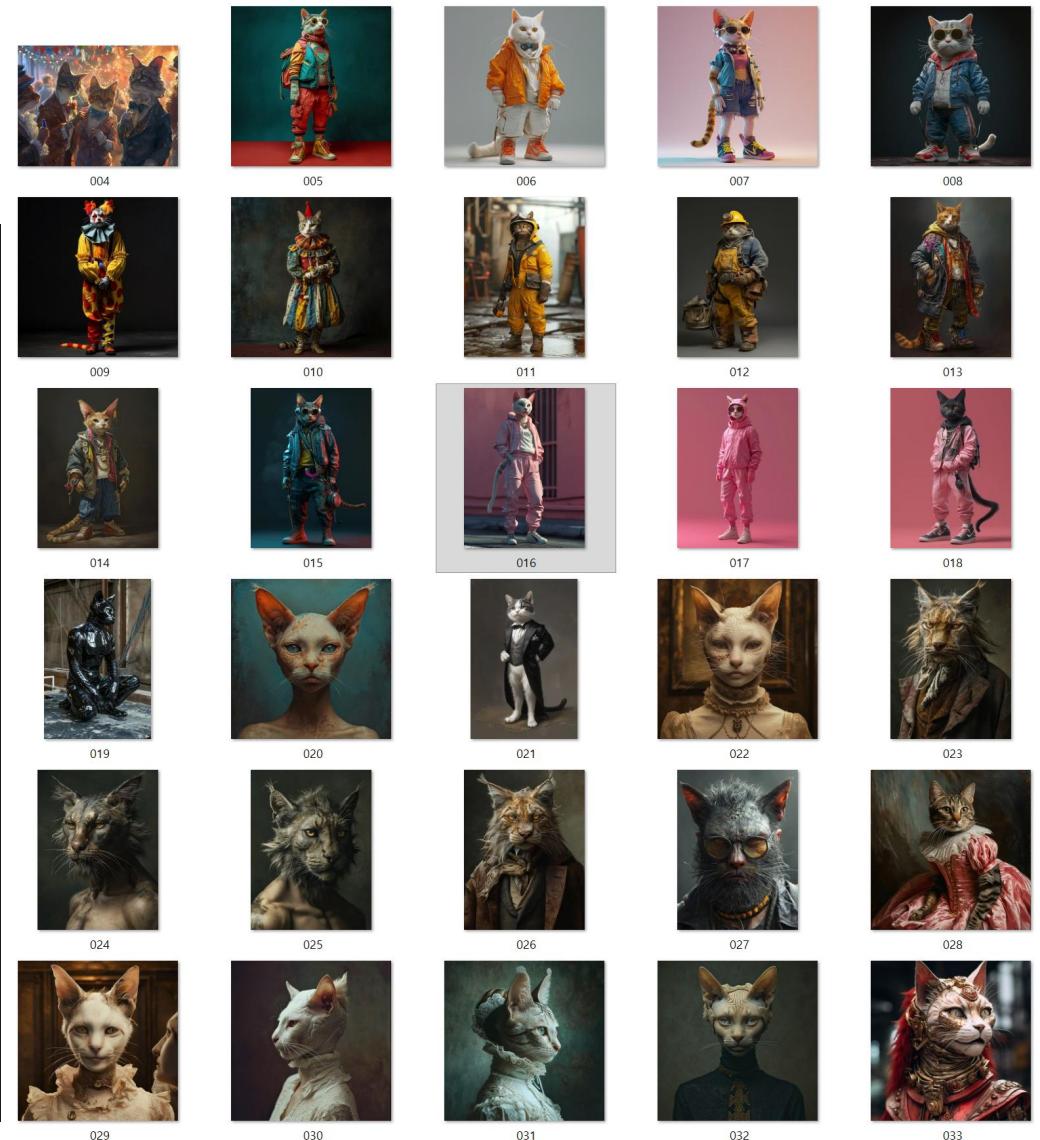
What did you think of this resource? [Upvote](#) [Downvote](#)

Details

| | |
|---------------|--------------------------|
| Type | LORA |
| Stats | 610 1.7K 5 |
| Reviews | Very Positive (54) |
| Published | Jan 4, 2024 |
| Base Model | SDXL 1.0 |
| Training | STEPS: 382 EPOCHS: 3 |
| Usage Tips | CLIP SKIP: 1 STRENGTH: 1 |
| Trigger Words | HUMEOW ⓘ |
| Hash | AUTOV2 3C4F40993E |
| AIR | civitai: 252637 @ 285058 |

1 File

About this version



<https://selgyun.tistory.com/entry/Stable-Diffusion-윈도우-설치하는-방법-SD-1>

<https://selgyun.tistory.com/entry/Stable-Diffusion-체크-포인트Checkpoint-학습-모델-추가하기-SD-2>

<https://selgyun.tistory.com/entry/Stable-Diffusion-기초-사용법-총정리-txt2img-SD-3>

[https://selgyun.tistory.com/6 Stable Diffusion 기초 사용법 총정리 : img2img - \[SD - 4\]](https://selgyun.tistory.com/6 Stable Diffusion 기초 사용법 총정리 : img2img - [SD - 4])

[https://selgyun.tistory.com/7 Stable Diffusion 로라\(Lora\) 설치 및 적용하는 법 - \[SD - 5\]](https://selgyun.tistory.com/7 Stable Diffusion 로라(Lora) 설치 및 적용하는 법 - [SD - 5])

[https://selgyun.tistory.com/8 Stable Diffusion VAE 적용하는 법 - \[SD - 6\]](https://selgyun.tistory.com/8 Stable Diffusion VAE 적용하는 법 - [SD - 6])

<https://selgyun.tistory.com/entry/Stable-Diffusion-임베딩-적용하는-법-SD-7>

<https://selgyun.tistory.com/entry/Stable-Diffusion-이미지-배경-삭제하는-법-SD-8>

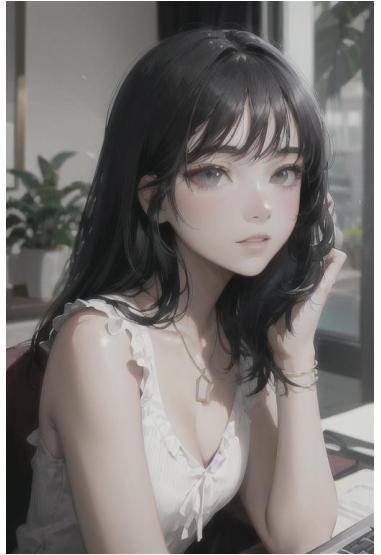
[https://selgyun.tistory.com/entry/Stable Diffusion 원하는 자세로 이미지 생성하기 - \[SD - 9\]](https://selgyun.tistory.com/entry/Stable Diffusion 원하는 자세로 이미지 생성하기 - [SD - 9])

[https://selgyun.tistory.com/entry/Stable Diffusion 태그 키워드 모음 - \[SD - 10\]](https://selgyun.tistory.com/entry/Stable Diffusion 태그 키워드 모음 - [SD - 10])

SD WebUI

Generation 임베딩 하이퍼네트워크 체크포인트 로라

img2img 스케치 인페인트 인페인트 스케치 인페인트 블러오기 배치



<https://civitai.com/models>

CIVITAI Pro

Models Search Civitai Create

Early Access ALL CHARACTER STYLE CONCEPT CLOTHING BASE MODEL POSES BACKGROUND TOOL VEHICLE BUILDINGS OBJECTS ASSETS ANIMAL ACTION

WAI-NSFW Illustrious SDXL Perfect Illustrious XL JANKU v5 NSFW Trained + NoobAI + RouWei Illustrious XL CyberRealistic Pony iLustMix Nova Anime XL Lazy Embeddings Lazy Positive Nova Furry XL Velvet's Mythic Fantasy Styles Flux + Pony

여기로 이미지 업로드 img2img 스케치 인페인트 인페인트 스케치

Generation Data

Prompt **TXT2IMG**
(best-quality:0.8), (best-quality:0.8), perfect anime illustration of a Deserted Groovy [Mooneye:megan fox:5], from inside of a Velvet Acadia National Park, Sun in the sky, shallow depth of field, Detailed illustration, Dark, Destructive art, Indirect light, F/2.8, Kodak portra, Spirals, opulent, adobe lightroom, photolab

Negative prompt
(worst quality:0.8), verybadimagegenegative_v1.3 easynegative, (surreal:0.8), (modernism:0.8), (art deco:0.8), (art nouveau:0.8)

Sampler DPM++ 2M Karras

Model v3

CFG scale 10 Steps 20

Seed 3939324605

로라(Lora)는 기존 학습 모델(체크포인트)에 새로운 이미지를 학습시킨 모델 파일입니다. 기존 학습 모델에 본인이 적용하고 싶은 스타일을 추가로 학습시킨 파일로, 프롬프트에 적용 시켜 좀 더 원하는 이미지를 생성할 수 있게 해 줍니다. Lora는 본인이 직접 만들 수도 있고, 사이트에서 다운로드하여 이용할 수도 있어요. 만약 게임을 제작하고 있다면, 주인공 캐릭터의 로라를 만들어 여러 주인공 이미지들을 생성할 수 있습니다. 이번에는 로라를 다운로드하여서 설명하겠습니다.

로라의 가중치는 0.1 ~ 1.0까지 반영할 수 있어요. 지금은 1로 들어가 있는데요. 지금은 로라를 100% 적용한다는 의미입니다. 그럼 0부터 각 단계별로의 이미지 변화를 보여 드리겠습니다. VAE는 이미지를 보정해주는 역할을 합니다. 만약 생성된 이미지가 흐리거나 퀄리티를 높이고 싶다면 VAE를 적용하면 됩니다. VAE도 스타일에 따라서 선택을 해주셔야 하는데요. 크게 2D 스타일(애니메이션 풍)과 실사 이미지 중에서 선택을 해주시면 됩니다.



The screenshot shows the Civitai website interface. At the top, there's a navigation bar with links like 'Home', 'Models' (which is selected), 'Images', 'Videos', 'Posts', 'Articles', 'Bounties', 'Challenges', 'Events', 'Updates', 'Shop', 'Create', and 'Sign In'. Below the navigation is a search bar and a sidebar with various links. The main content area displays two images of a chibi-style girl: one playing a guitar outdoors and another sitting on a couch indoors. To the right of these images is a detailed card for the 'blindbox_v1_mix' version of the model. The card includes a 'Bid' button, a 'Download (144.11 MB)' button, and a 'Verified: 2 years ago' badge. It also lists the following details:

| | |
|---------------|---------------------------------|
| Type | LORA |
| Stats | 228,833 downloads, 333 reviews |
| Reviews | Overwhelmingly Positive (4.543) |
| Published | Apr 1, 2023 |
| Base Model | SD 1.5 |
| Trigger Words | FULL BODY, CHIBI |
| Hash | AUTOV2 525491E628 |

Below the card, there's a section for 'About this version' with statistics: 957 comments and 14.5K likes. A user profile for 'samecorner' is shown, indicating they joined on Feb 15, 2023, and a 'Follow' button is available. At the bottom of the page, there's a footer with links to 'Terms of Service', 'Privacy', 'Safety', 'Newsroom', 'API', 'Status', 'Wiki', 'Education', 'Creators', 'Careers', '18 U.S.C. §2257', and a 'Support' link.

BlindBox Artstyle 更新了V3版本，调整了眼睛，俗话说眼睛是心灵的窗口。但是目前还是实验性质的，从稳定性上还是推荐v1_mix版本。

Updated v3 version, adjusted the eyes, as the saying goes, the eyes are the windows to the soul.

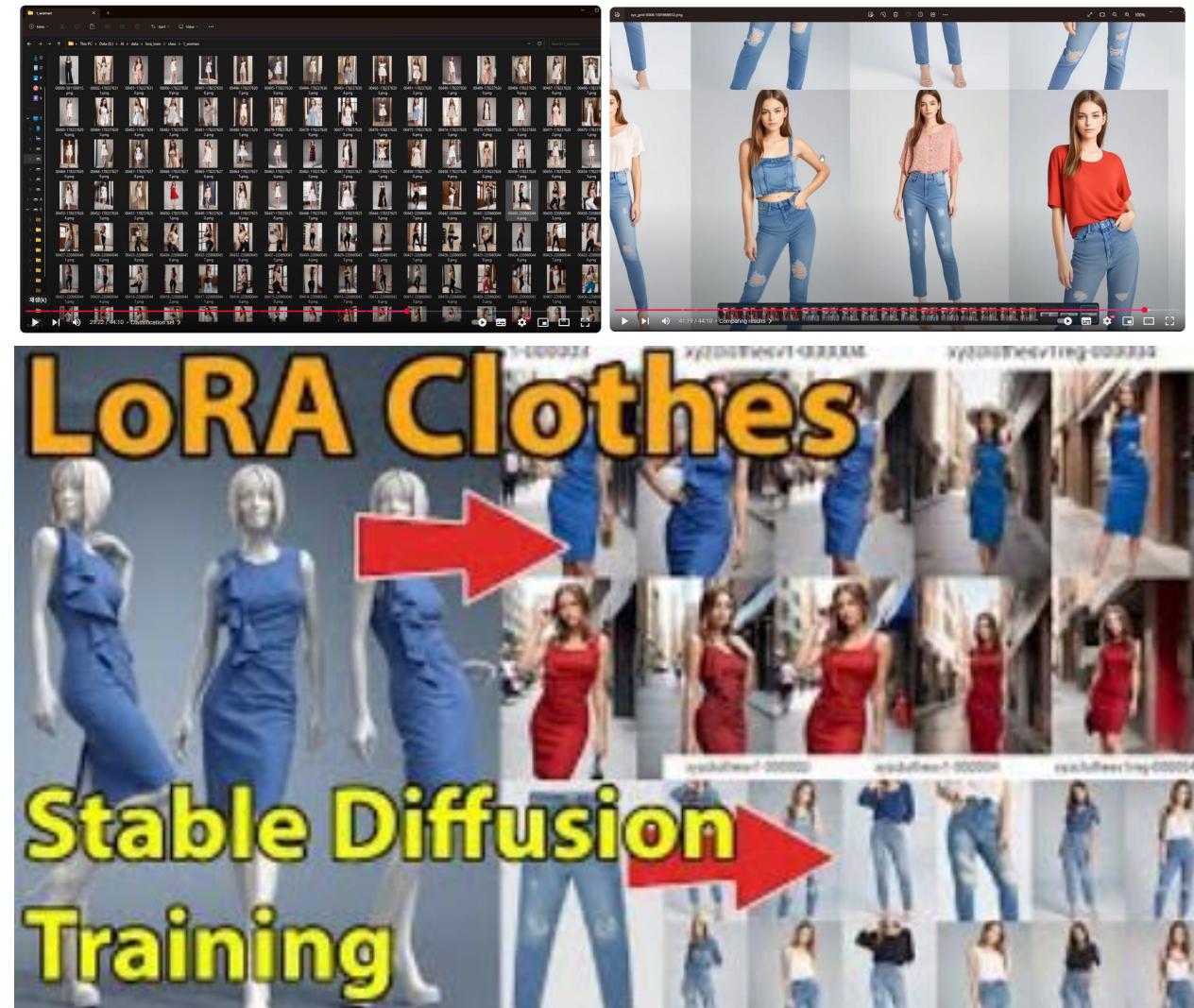
But it is still experimental, from the stability or recommended v1_mix version.

LoRA Clothes

this video discusses a possible method for training clothes in stable diffusion using kohya ss LoRA models...which can be useful for fashion business/cloth stores or other businesses or users. the goal here is to train objects such as clothes, use existing clothes which are on plastic mannequins or without to dress them on new real person (or AI generated person) in different poses/backgrounds/settings it is possible to train clothes without having them on mannequins or existing models using the cloth pictures only but results are not as good as here, with mannequins it gives better results and we can automatically remove the mannequin from the generated images with help of captioning and some proper data set preparation.

the video will explain multiple subject training, object training method in general and compare results from regularized and non regularized sets

- 1- we can use 1 to 20 images for training, the more the better
- 2- clothes should ideally be seen from different angles and how they are dressed on a plastic mannequin or another real person
- 3- the face of the mannequin or person must not repeat in most images to avoid learning it too
- 4- we can change the cloth colors afterwards using prompts in most cases
- 5- regularization can make model more flexible but reduce resemblance slightly
- 6- training multiple subjects is possible, but can be annoying and less flexible than single subject because each subject may require different number of training steps
- 7- use after detailer to get good full boy faces in all your training targets.
- 8- it is possible to generate more than one prompt in Kohya ss by putting one prompt in each line in the sample image generation box



Stable Diffusion 파인 튜닝은 사전 학습된 모델을 특정 작업이나 데이터셋에 맞추기 위해 추가 학습하는 과정입니다. Stable Diffusion의 파인 튜닝은 다양한 방법으로 수행될 수 있으며, DreamBooth와 LoRA 같은 대표적인 기법을 소개하겠습니다.

DreamBooth는 Google Research에서 개발한 기술로, 사용자 지정 데이터셋을 사용하여 사전 학습된 모델을 특정 요구 사항에 맞추도록 조정합니다. DreamBooth는 이미지 생성 모델, 특히 텍스트-이미지 변환 모델에 자주 사용됩니다.

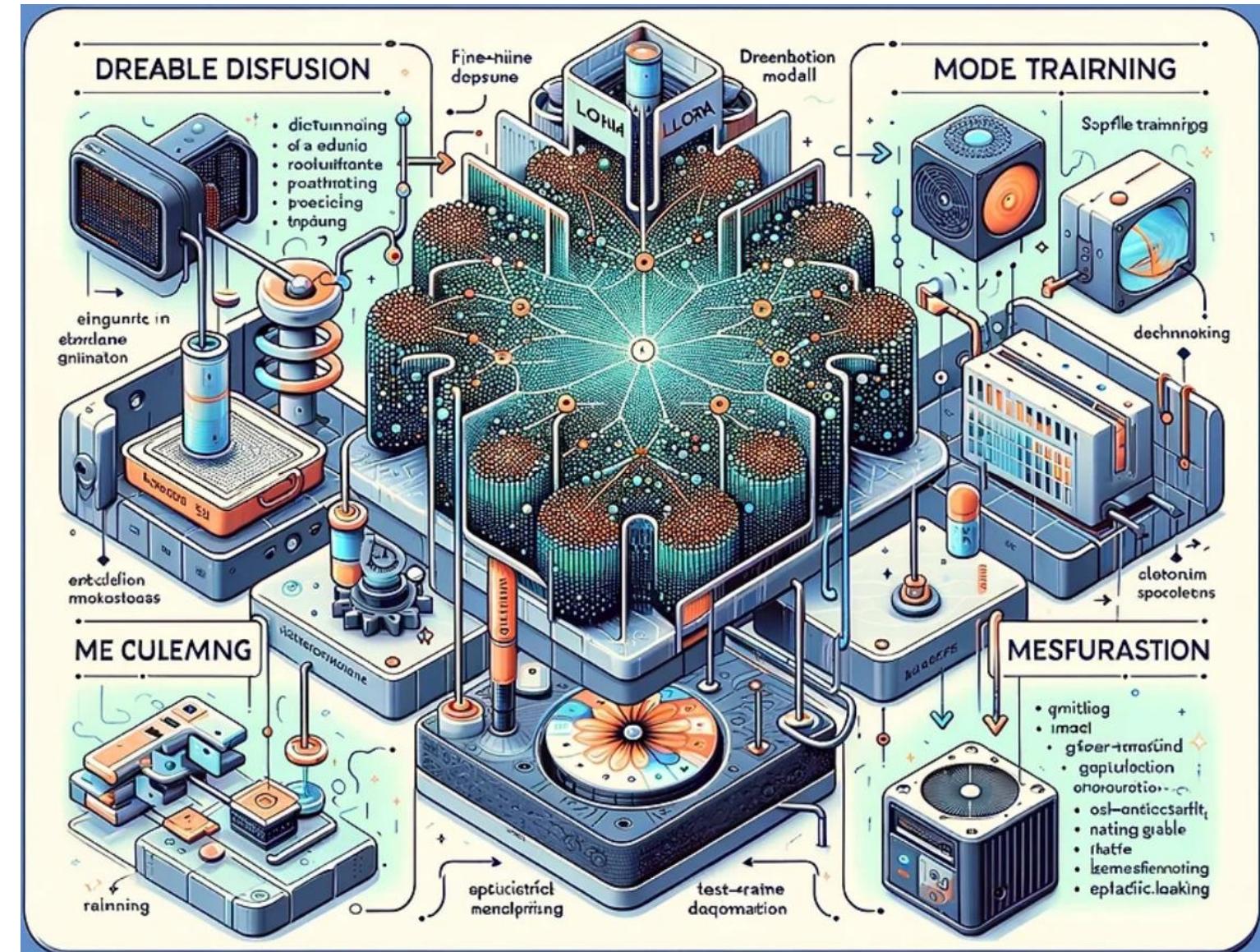
- 세부 조정: 특정한 소수의 이미지를 사용하여 모델을 파인 튜닝할 수 있습니다.
- 높은 품질: 입력 데이터의 세밀한 특징을 학습하여 높은 품질의 출력을 생성합니다.
- 전용 모델 생성: 특정 인물이나 스타일을 학습하여, 유사한 결과물을 일관되게 생성할 수 있습니다.

LoRA(Low-Rank Adaptation)는 사전 학습된 모델의 효율적인 파인 튜닝을 위해 사용되는 방법입니다. 주로 텍스트 생성 모델에서 사용되며, 모델의 특정 부분을 저랭크 행렬로 대체하는 방식으로 작동합니다.

- 효율성: 모델의 전체 가중치를 업데이트하는 대신, 저랭크 행렬만 업데이트하여 메모리와 계산 효율성을 높입니다.
- 빠른 학습: 기존 모델의 대부분을 그대로 유지하면서 필요한 부분만 수정하므로 빠른 학습이 가능합니다.
- 적은 데이터 요구: 적은 양의 데이터로도 효과적인 파인 튜닝이 가능합니다.

Stable Diffusion 파인 튜닝의 응용 사례

- 광고 디자인: 특정 브랜드의 이미지와 스타일에 맞춘 광고 이미지를 자동으로 생성.
- 게임 개발: 게임 캐릭터나 배경 이미지를 텍스트 설명을 통해 자동으로 생성.
- 예술 창작: 작가의 스타일에 맞춘 예술 작품을 생성.
- 교육 자료 제작: 교육 콘텐츠와 관련된 이미지를 자동으로 생성하여 시각적 학습 자료 제공.



인공지능(AI)과 딥러닝은 현대 기술의 중심에 있으며, 이미지 생성 모델은 그 중에서도 많은 주목을 받고 있습니다. **Stable Diffusion**은 최근 많은 관심을 받고 있는 이미지 생성 모델 중 하나로, 텍스트를 기반으로 고품질 이미지를 생성하는 데 사용됩니다. 이 블로그 글에서는 **Stable Diffusion**의 개념과 이를 파인 튜닝하는 100 양한 방법을 소개합니다.

1. DreamBooth란 무엇인가?

DreamBooth는 이미지 생성 및 변형을 위해 딥러닝 모델을 훈련시키는 기법입니다. 이 기법은 특히 특정 스타일이나 특성을 가진 이미지를 학습하고, 이를 바탕으로 새로운 이미지를 생성하는 데 유용합니다. 이를 통해 예술가, 디자이너, 연구자들은 기존의 이미지를 변형하거나 새로운 스타일의 이미지를 쉽게 생성할 수 있습니다.



스타일 Fine-tuning 예시

2. DreamBooth의 주요 기능

- 스타일 전이: 특정 스타일의 이미지를 학습하여 새로운 이미지에 그 스타일을 적용.
- 이미지 변형: 입력 이미지의 특성을 바탕으로 새로운 이미지를 생성.
- 고해상도 이미지 생성: 기존 모델보다 더 높은 해상도의 이미지를 생성.

3. 필요한 라이브러리 설치

DreamBooth를 사용하려면 몇 가지 필수 라이브러리를 설치해야 합니다. 아래 명령어를 통해 필요 한 패키지를 설치할 수 있습니다.

```
pip install diffusers transformers accelerate
pip install torch torchvision torchaudio
pip install datasets
pip install wandb
```

4. 데이터셋 준비

DreamBooth를 활용하기 위해서는 훈련에 사용할 데이터셋이 필요합니다. 데이터셋은 다양한 스타일의 이미지를 포함해야 하며, 이를 통해 모델이 다양한 특성을 학습할 수 있습니다. 데이터셋은 로컬에 저장하거나, 공공 데이터셋을 사용할 수 있습니다.

5. 코드 설명 및 구현

이제 본격적으로 DreamBooth의 코드를 살펴보겠습니다. 아래 코드는 train_dreambooth.py 파일의 주요 부분을 설명합니다.

5.1. 라이브러리 임포트

```
import argparse
import logging
import os
import torch
from accelerate import Accelerator
from diffusers import StableDiffusionPipeline
from transformers import AutoTokenizer
from datasets import load_dataset
from torch.utils.data import DataLoader
```

여기서는 필요한 라이브러리를 임포트합니다. argparse는 명령행 인자를 처리하고, logging은 로그를 기록하는 데 사용됩니다. torch는 PyTorch 라이브러리로 딥러닝 모델을 구현하고 훈련하는 데 사용됩니다.

5.2. 하이퍼파라미터 설정

```
parser = argparse.ArgumentParser(description="Training script for DreamBooth")
parser.add_argument("--pretrained_model_name_or_path", type=str, required=True, help="Path to pretrained model")
parser.add_argument("--train_data_dir", type=str, required=True, help="Directory containing training data")
parser.add_argument("--output_dir", type=str, required=True, help="Directory to save the trained model")
parser.add_argument("--num_train_epochs", type=int, default=4, help="Number of training epochs")
parser.add_argument("--learning_rate", type=float, default=5e-5, help="Learning rate")
args = parser.parse_args()
```

여기서는 하이퍼파라미터를 설정합니다. pretrained_model_name_or_path는 사전 학습된 모델의 경로를 지정하고, train_data_dir은 훈련 데이터가 저장된 디렉토리를 지정합니다. output_dir은 훈련된 모델을 저장할 디렉토리를 지정합니다.

5.3. 데이터 로드 및 전처리

```
dataset = load_dataset("imagefolder", data_dir=args.train_data_dir)
train_dataloader = DataLoader(dataset['train'], batch_size=16, shuffle=True)
```

여기서는 데이터셋을 로드하고, 데이터 로더를 설정합니다. load_dataset 함수는 주어진 디렉토리에서 이미지를 로드하며, DataLoader는 배치 단위로 데이터를 불러오는 역할을 합니다.

5.4. 모델 초기화

```
model = StableDiffusionPipeline.from_pretrained(args.pretrained_model_name_or_path)
tokenizer = AutoTokenizer.from_pretrained(args.pretrained_model_name_or_path)
```

여기서는 사전 학습된 모델과 토크나이저를 초기화합니다. StableDiffusionPipeline은 이미지 생성 모델을 불러오고, AutoTokenizer는 텍스트 데이터를 토큰화하는 데 사용됩니다.

5.5. 훈련 루프

```
optimizer = torch.optim.AdamW(model.parameters(), lr=args.learning_rate)
accelerator = Accelerator()

model, optimizer, train_dataloader = accelerator.prepare(model, optimizer, train_dataloader)

for epoch in range(args.num_train_epochs):
    model.train()
    for batch in train_dataloader:
        outputs = model(**batch)
        loss = outputs.loss
        accelerator.backward(loss)
        optimizer.step()
        optimizer.zero_grad()

    print(f"Epoch {epoch} completed.")
    model.save_pretrained(args.output_dir)
```

여기서는 모델 훈련을 위한 루프를 설정합니다. optimizer는 AdamW 옵티마이저를 사용하여 모델의 파라미터를 최적화합니다. accelerator는 모델과 옵티마이저, 데이터로더를 가속화합니다. 각 에포크마다 모델을 훈련시키고, 손실을 계산하여 역전파를 수행합니다. 훈련이 끝나면 모델을 저장합니다.

5.6. 모델 저장 및 평가

훈련이 완료되면, 모델을 저장하고 평가합니다. 저장된 모델은 새로운 이미지 생성을 위해 사용할 수 있습니다.

6. DreamBooth를 활용한 실제 예제

위 코드를 이용하면 DreamBooth를 사용하여 다양한 스타일의 이미지를 생성해볼 수 있습니다. 예를 들어, 특정 예술가의 스타일을 학습한 모델을 통해 새로운 작품을 생성할 수 있습니다. 이를 위해 다양한 예술가의 작품을 데이터셋으로 사용하고, DreamBooth를 활용하여 모델을 훈련합니다.

7. 최적화 및 튜닝

모델의 성능을 최적화하기 위해 하이퍼파라미터를 조정하고, 더 많은 데이터를 활용할 수 있습니다.

또한, 모델의 구조를 변경하거나, 추가적인 정규화 기법을 도입하여 성능을 향상시킬 수 있습니다.

IP-Adapter란 무엇인가?

IP-Adapter는 이미지 프롬프트 어댑터로, 기존 모델을 수정하지 않고도 이미지 프롬프트를 사용할 수 있게 해줍니다. 이는 텍스트와 이미지 특징을 구분하여 별도의 크로스-어텐션 레이어를 추가하는 메커니즘을 통해 이루어집니다. 이러한 구조는 모델이 보다 이미지 특화된 특징을 학습할 수 있도록 도와줍니다.



파인튜닝을 통한 이미지 제작

얼굴 모델 생성

얼굴 이미지를 생성하는 것은 매우 어려운 작업입니다. 이를 위해 특화된 IP-Adapter 체크포인트를 사용할 수 있습니다.

```
pipeline.load_ip_adapter("h94/IP-Adapter", subfolder="models", weight_name="ip-adapter-full-face_sd15.bin")
pipeline.set_ip_adapter_scale(0.5)
```

```
image = load_image("https://huggingface.co/datasets/huggingface/documentation-images/resolve/main/diffusers/ip_adapter_einstein_base.png")
generator = torch.Generator(device="cpu").manual_seed(26)
```

```
image = pipeline(
    prompt="A photo of Einstein as a chef, wearing an apron, cooking in a French restaurant",
    ip_adapter_image=image,
    negative_prompt="lowres, bad anatomy, worst quality, low quality",
    num_inference_steps=100,
    generator=generator,
).images[0]
image
```

멀티 IP-Adapter

여러 IP-Adapter를 동시에 사용하여 특정 스타일의 이미지를 생성할 수 있습니다.

```
pipeline.load_ip_adapter(
    "h94/IP-Adapter",
    subfolder="sdxl_models",
    weight_name=[("ip-adapter-plus_sdxl_vit-h.safetensors", "ip-adapter-plus-face_sdxl_vit-h.safetensors")]
)
pipeline.set_ip_adapter_scale([0.7, 0.3])
pipeline.enable_model_cpu_offload()
```

```
face_image = load_image("https://huggingface.co/datasets/YiYiXu/testing-images/resolve/main/women_input.png")
style_folder = "https://huggingface.co/datasets/YiYiXu/testing-images/resolve/main/style_ziggy"
style_images = [load_image(f"{style_folder}/img{i}.png") for i in range(10)]
```

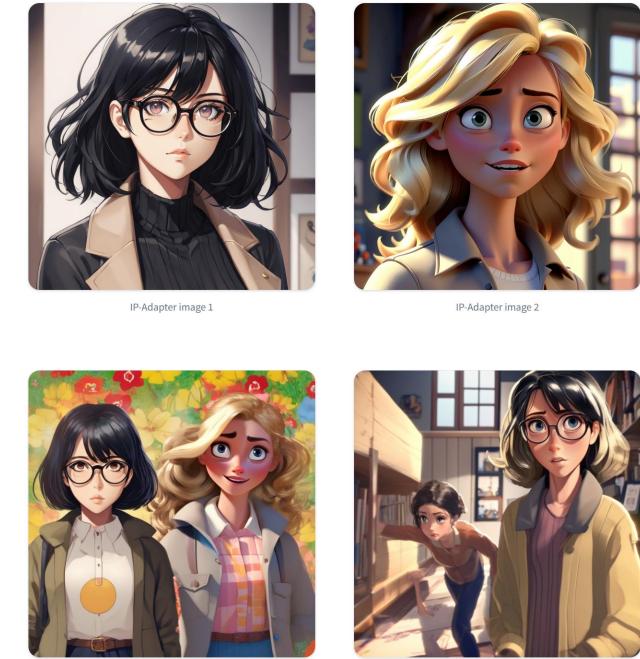
```
generator = torch.Generator(device="cpu").manual_seed(0)

image = pipeline(
    prompt="wonderwoman",
    ip_adapter_image=[style_images, face_image],
    negative_prompt="monochrome, lowres, bad anatomy, worst quality, low quality",
    num_inference_steps=50, num_images_per_prompt=1,
    generator=generator,
).images[0]
image
```

결론

IP-Adapter는 Stable Diffusion 모델의 성능을 극대화할 수 있는 강력한 도구입니다. 이를 통해 모델의 이미지 생성 능력을 향상시키고, 보다 정확하고 다양한 이미지를 생성할 수 있습니다. 이 글에서 소개한 방법들을 활용하여 여러분의 프로젝트에 IP-Adapter를 적용해 보세요.

추가 정보는 [Hugging Face의 공식 문서](#)에서 확인할 수 있습니다.



LoRA

LoRA (Low-Rank Adaptation) is a method for quickly training a model for a new task. It works by freezing the original model weights and adding a small number of *new* trainable parameters. This means it is significantly faster and cheaper to adapt an existing model to new tasks, such as generating images in a new style.

LoRA checkpoints are typically only a couple hundred MBs in size, so they're very lightweight and easy to store. Load these smaller set of weights into an existing base model with load_lora_weights() and specify the file name.

The screenshot shows a browser window displaying the Hugging Face documentation for LoRA. The URL in the address bar is https://huggingface.co/docs/diffusers/main/en/tutorials/using_peft_for_inference. The page content is as follows:

You are viewing *main* version, which requires [installation from source](#). If you'd like regular pip install, checkout the latest stable version ([v0.35.1](#)).

Join the Hugging Face community
and get access to the augmented documentation experience

LoRA

LoRA (Low-Rank Adaptation) is a method for quickly training a model for a new task. It works by freezing the original model weights and adding a small number of *new* trainable parameters. This means it is significantly faster and cheaper to adapt an existing model to new tasks, such as generating images in a new style.

LoRA checkpoints are typically only a couple hundred MBs in size, so they're very lightweight and easy to store. Load these smaller set of weights into an existing base model with load_lora_weights() and specify the file name.

text-to-image text-to-video

```
import torch
from diffusers import AutoPipelineForText2Image

pipeline = AutoPipelineForText2Image.from_pretrained(
    "stabilityai/stable-diffusion-xl-base-1.0",
    torch_dtype=torch.float16
).to("cuda")
pipeline.load_lora_weights(
    "ostris/super-cereal-sdxl-lora",
    weight_name="cereal_box_sdxl_v1.safetensors",
    adapter_name="cereal"
)
pipeline("bears, pizza bites").images[0]
```

The load_lora_weights() method is the preferred way to load LoRA weights into the UNet and text encoder because it can handle cases where:

- the LoRA weights don't have separate UNet and text encoder identifiers

On the right side of the page, there is a sidebar with a tree view of the documentation structure:

- LoRA**
 - torch.compile
 - Weight scale
 - Scale scheduling
 - Hotsplitting
 - Compiled models
 - Merge
 - set_adapters
 - add_weighted_adapter
 - fuse_lora
 - Manage
 - set_adapters
 - save_lora_adapter
 - unload_lora_weights
 - disable_lora
 - get_active_adapters
 - get_list_adapters
 - delete_adapters
 - Resources