



Perspectives on text-to-image models



Dennis Wilson
2022



TECH

Why Silicon Valley is so excited about awkward drawings done by artificial intelligence

PUBLISHED SAT, OCT 8 2022-8:00 AM EDT



Kif Leswing
@KIFLESWING

Stability AI, the startup behind Stable Diffusion, raises \$101M

DALL-E 2, the future of AI research, and OpenAI's business model



COSMOPOLITAN

the A.I. issue

Meet the World's First Artificially Intelligent Magazine Cover



Art Made by AI Wins Fine Arts Competition

AI-generated artwork won a recent art competition in the US, sparking controversy and fury among artists

by Belinda Teoh — September 13, 2022 in Art, Culture, Society, Tech



TEXT DESCRIPTION

An astronaut Teddy bears A bowl of
soup

riding a horse lounging in a tropical resort
in space playing basketball with cats in
space

as a children's book illustration in a
minimalist style in a watercolor style



DALL-E 2



Are text-to-image models

- The future of art?
- The end of art?
- The future of AI?
- Legal?
- Just an example of Silicon hype?

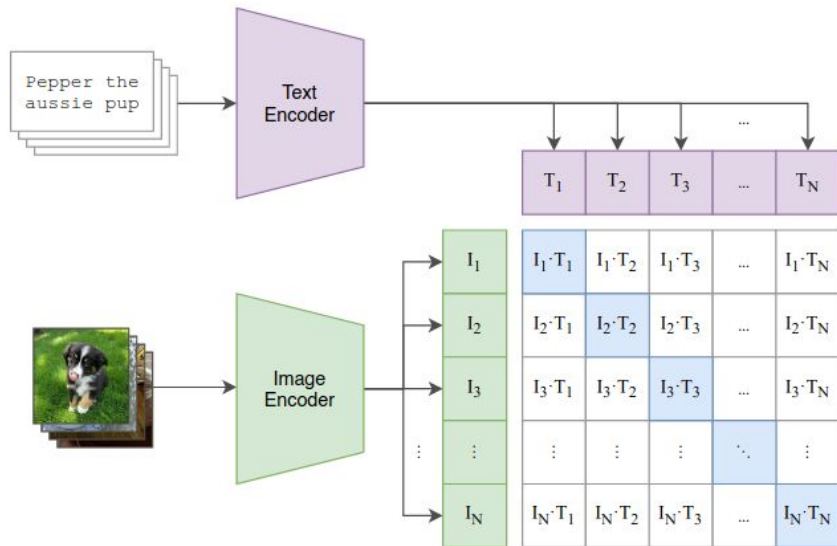
Collapsed Crypto Exchange FTX Could Owe More Than 1 Million Creditors

In a court filing on Monday, the cryptocurrency exchange said it was in touch with “dozens” of regulators across the world.

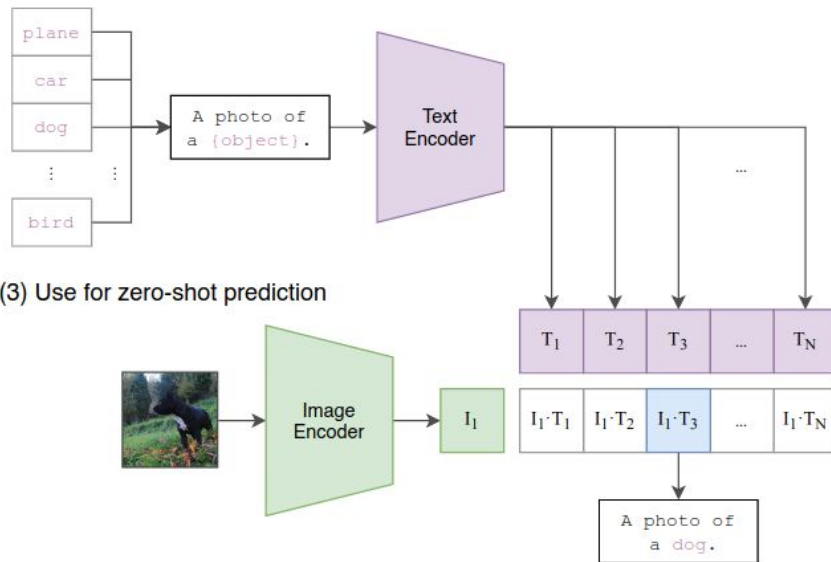


Technical overview: CLIP

(1) Contrastive pre-training

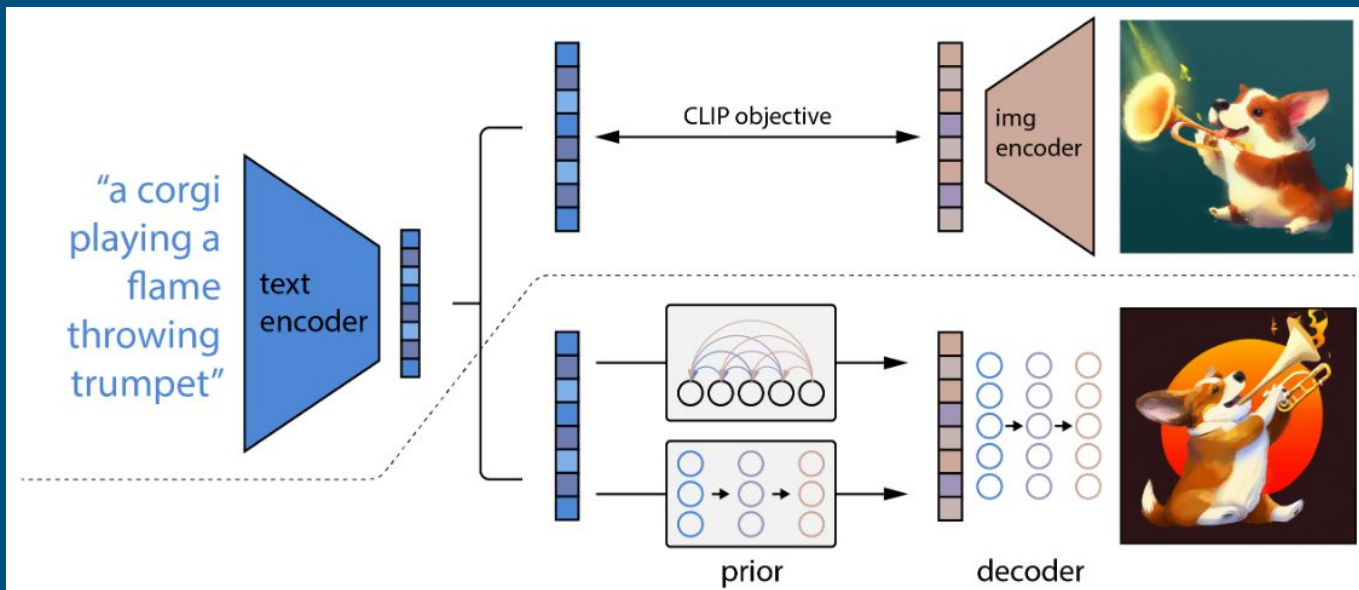


(2) Create dataset classifier from label text



Radford, Alec, et al. "Learning transferable visual models from natural language supervision." *International Conference on Machine Learning*. PMLR, 2021.

Technical overview: unCLIP



Ramesh, Aditya, et al. "Hierarchical text-conditional image generation with clip latents."
arXiv preprint arXiv:2204.06125 (2022).

Technical overview: Training data

LAION-5B: A NEW ERA OF OPEN LARGE-SCALE MULTI- MODAL DATASETS

by: Romain Beaumont, 11 Nov, 2022

We present a dataset of 5,85 billion CLIP-filtered image-text pairs, 14x bigger than LAION-400M, previously the biggest openly accessible image-text dataset in the world.

Authors: Christoph Schuhmann, Richard Vencu, Romain Beaumont, Theo Coombes, Cade Gordon, Aarush Katta, Robert Kaczmarczyk, Jenia Jitsev

Technical overview: Training data

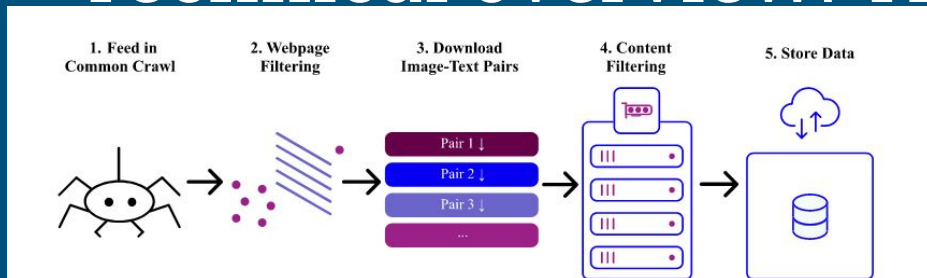


Figure 2: **Overview of the acquisition pipeline:** Files are downloaded, tracked, and undergo distributed inference to determine inclusion. Those above the specified CLIP threshold are saved.

Q: An armchair that looks like an apple



C: Green Apple Chair

Q: A dog rolling in the snow at sunset



C: sun snow dog

Q: A graphic design color palette



C: Color Palettes

Q: pink photo of Tokyo



C: pink, japan, aesthetic image

Figure 3: **LAION-5B examples.** Sample images from a nearest neighbor search in LAION-5B using CLIP embeddings. The image and caption (C) are the first results for the query (Q).

Schuhmann, Christoph, et al. "LAION-5B: An open large-scale dataset for training next generation image-text models." arXiv preprint arXiv:2210.08402 (2022).

RGPD violations in LAION-5B

7. Your rights

If your personal data is processed on the basis of consent which you have given us, you have the right to revoke your consent at any time. The revocation of consent does not affect the legality of the processing performed on the basis of the consent until the time of revocation. You can contact us at any time to exercise your right to revoke consent.

7.2 Right to confirmation

You have the right to request confirmation from the controller that we are processing personal data concerning you. You can request this confirmation at any time using the contact details above.

Technical overview: Training data

In the current form, we consider this dataset a research artefact and strongly advocate **academic use-only** and advise careful investigation of downstream model biases (Appendix Sec. [G.2](#)). Additionally, we encourage users to use the described tools and to transparently explore and, subsequently, report further not yet detected content and model behaviour to our dataset repository¹⁴, and help to further advance existing approaches for data curation using the real-world large dataset introduced here.

Privacy. We comment on privacy issues arising from Common Crawl as source of links in LAION-5B and measures undertaken to handle those in the Appendix Sec. [G.1](#)

Schuhmann, Christoph, et al. "LAION-5B: An open large-scale dataset for training next generation image-text models." arXiv preprint arXiv:2210.08402 (2022).

But...

DALL-E

Pricing

Simple and flexible. Only pay for what you use.

[GET STARTED](#)[CONTACT SALES](#)

Image models

Build DALL-E directly into your apps to generate and edit novel images and art. Our image models offer three tiers of resolution for flexibility.

[LEARN MORE >](#)

RESOLUTION

1024×1024

512×512

256×256

PRICE

\$0.020 / image

\$0.018 / image

\$0.016 / image

Midjourney

	Free trial	Basic	Standard	Corporate***
Price	Free	\$10 / month	\$30 / month	\$600 / year
Fast GPU time	25 min* / lifetime	200 min* / month	15 hrs* / month	120 hrs* / year
Relax GPU time	No	No	Unlimited	Unlimited
Metered mode**	No	Yes	Yes	Yes
Personal Bot Chat	No	Yes	Yes	Yes
Private visibility	No	for +\$20 / month	for +\$20 / month	Yes

* 1 generation job roughly takes 1 gpu minute, upscales are longer, variants are quicker

** Metered mode can be enabled to use additional fast time for \$4 / gpu hour

*** Corporate plan is required for employees of companies with revenues over \$1Million / year

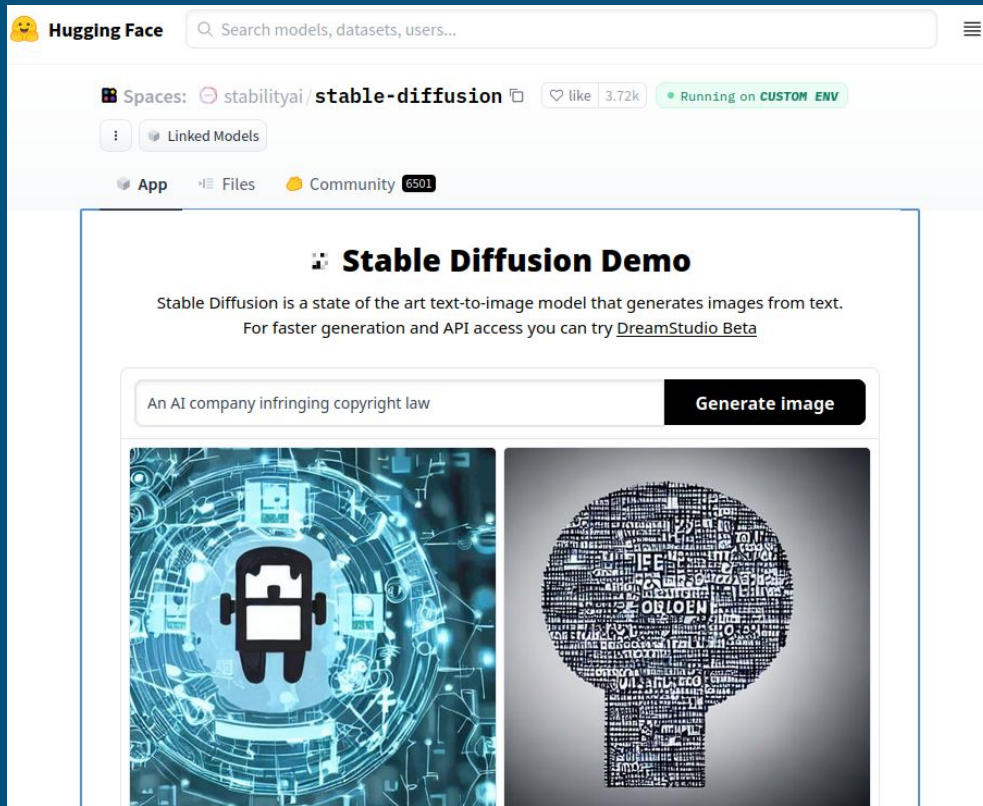
Stable Diffusion

Trained on LAION-5B

Public model release

Web interfaces for free

Prospects for inclusion in photo-editing and generation tools



Brief overview of copyright law (US-based)

- With the stated purpose to promote art and culture, copyright law assigns a set of exclusive rights to authors: to make and sell copies of their works, to create derivative works, and to perform or display their works publicly.
- The public domain (PD) consists of all the creative work to which no exclusive intellectual property rights apply.
- Fair use is a doctrine in United States law that permits limited use of copyrighted material without having to first acquire permission from the copyright holder.

Fair Use

The fair use of a copyrighted work, including such use by reproduction in copies or phonorecords or by any other means specified by that section, for purposes such as criticism, comment, news reporting, teaching (including multiple copies for classroom use), scholarship, or research, is not an infringement of copyright.

17 USC Section 107

Is text-to-image fair use?

8/16/22

Has anyone had any luck getting Midjourney or Stable Diffusion to output RJ Palmer's work? I'm trying to make some realistic Pokémon AI art in the Arvalis style

2

Here's an attempt at a Charizard. Prompt was "realistic Charizard, RJ Palmer, trending on artstation, trending on deviantart"

RJ Palmer @arvalis · Aug 31

Replying to @arvalis

It really captured my Charizard aesthetic as you can see

Is text-to-image fair use?

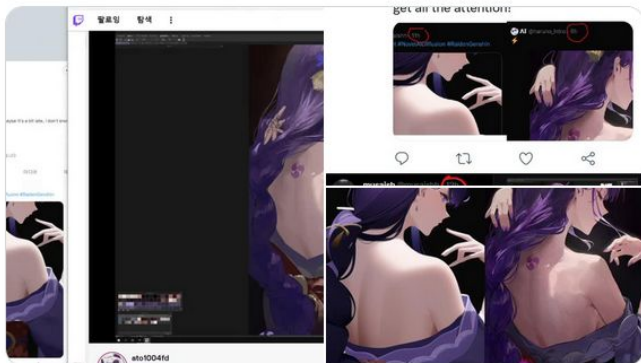


Genel Jumalon 🌸 Holiday Matsuri
@GenelJumalon

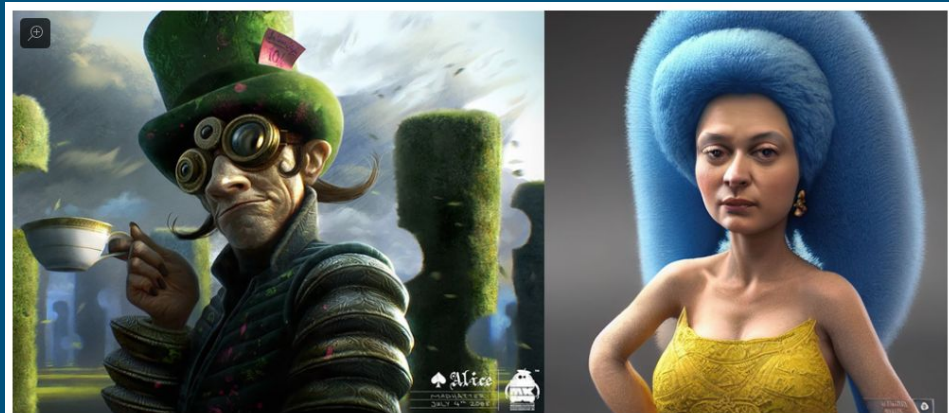
During a Twitch stream AT (@haruno_intro) had their art stolen.

The thief then finished the sketch by using NovelAI and posted on their Twitter before AT finish it.

Then had the AUDACITY to demand a "proper reference" from them.



5:32 AM · Oct 13, 2022



On the left is a piece by award-winning Hollywood artist Michael Kutsche, while on the right is a piece of AI art that's claimed to have copied his iconic style, including a blurred, incomplete signature

Image: Michael Kutsche

<https://kotaku.com/ai-art-dall-e-midjourney-stable-diffusion-copyright-1849388060>

Training ML models on copyrighted data

- Authors Guild, Inc. v. Google, Inc.
- Training of Google Books recommendation models on book contents (copyrighted information)
- First decision (2013): *In my view, Google Books provides significant public benefits. It advances the progress of the arts and sciences, while maintaining respectful consideration for the rights of authors and other creative individuals, and without adversely impacting the rights of copyright holders.*
- Second decision (2014): *Google's unauthorized digitizing of copyright-protected works, creation of a search functionality, and display of snippets from those works are non-infringing fair uses. The purpose of the copying is highly transformative, the public display of text is limited, and the revelations do not provide a significant market substitute for the protected aspects of the originals. Google's commercial nature and profit motivation do not justify denial of fair use.*

Content generated by ML models

- *The crucial question appears to be whether the “work” is basically one of **human authorship**, with the computer merely being an assisting instrument, or whether the traditional element of authorship in the work (literary, artistic or musical expression or elements of selection, arrangements, etc.) were actually conceived and executed not by man but by a machine.*
- <https://www.copyright.gov/rulings-filings/review-board/docs/a-recent-entrance-to-paradise.pdf>
- The Copyright Act of 1976

A story of a different generative model: Github Copilot

Your AI pair programmer

GitHub Copilot uses the OpenAI Codex to suggest code and entire functions in real-time, right from your editor.

[Explore docs >](#)

```
ts sentiments.ts  write_sql.go  parse_expenses.py  addresses.rb

1 #!/usr/bin/env ts-node
2
3 import { fetch } from "fetch-h2";
4
5 // Determine whether the sentiment of text is positive
6 // Use a web service
7 async function isPositive(text: string): Promise<boolean> {
8   const response = await fetch('http://text-processing.com/api/sentiment/', {
9     method: "POST",
10    body: `text=${text}`,
11    headers: {
12      "Content-Type": "application/x-www-form-urlencoded",
13    },
14  });
15  const json = await response.json();
16  return json.label === "pos";
17 }
```

Copilot

Is Copilot fair use?

The image shows a Visual Studio Code editor window with the title bar "GitHub Copilot - javastuff - Visual Studio Code". The editor has three tabs: "Untitled-1.cpp", "void cs_gaxpy (const cs *A, const double", and "sparse matrix transpose Untitled-2". The active tab is "sparse matrix transpose Untitled-2", which contains the following C code:

```
1 // sparse matrix transpose
2 cs_
```

Below the editor is a terminal window with the following output:

```
File Edit View Search Terminal Help
#include "cs.h"
/* C = A' */
cs *cs_transpose (const cs *A, csi values)
{
    csi p, q, j, *Cp, *Ci, n, m, *Ap, *Ai, *w ;
    double *Cx, *Ax ;
    cs *C ;
    if (!CS_CSC (A)) return (NULL) ; /* check inputs */
    m = A->m ; n = A->n ; Ap = A->p ; Ai = A->i ; Ax = A->x ;
    C = cs_spalloc (n, m, Ap [n], values && Ax, 0) ; /* allocate result */
    w = cs_calloc (m, sizeof (csi)) ; /* get workspace */
    if (!C || !w) return (cs_done (C, w, NULL, 0)) ; /* out of memory */
    Cp = C->p ; Ci = C->i ; Cx = C->x ;
    for (p = 0 ; p < Ap [n] ; p++) w [Ai [p]]++ ; /* row counts */
    cs_cumsum (Cp, w, m) ; /* row pointers */
    for (j = 0 ; j < n ; j++)
    {
        for (p = Ap [j] ; p < Ap [j+1] ; p++)
        {
            Ci [q = w [Ai [p]]++] = j ; /* place A(i,j) as entry C(j,i) */
            if (Cx) Cx [q] = Ax [p] ;
        }
    }
    return (cs_done (C, w, NULL, 1)) ; /* success; free w and return C */
}
```

On the right side of the editor, there is a panel titled "GitHub Copilot" showing a list of solutions for the current code snippet. The first solution is "Synthesizing 10/10 solutions" and is expanded, showing the following C code:

```
1 Synthesizing 10/10 solutions
2 =====
3
4 Accept Solution
5 cs_transpose (const cs *A, int values)
6 {
7     int p, q, j, *Cp, *Ci, *w, *Ti, *Tj ;
8     double *Cx, *Tx ;
9     cs *T ;
10    if (!CS_CSC (A) || !CS_TRIPLET (A)) return (NULL) ; // check inputs
11    T = cs_spalloc (A->n, A->m, A->nzmax, values && A->x, 0) ;
12    w = cs_calloc (A->m, sizeof (int)) ; // get workspace
13    if (!T || !w) return (cs_done (T, NULL, w, 0)) ; // out of memory
14    Cp = T->p ; Ci = T->i ; Cx = T->x ; Ti = A->i ; Tj = A->j ;
15    for (p = 0 ; p < A->nzmax ; p++) w [Ti [p]]++ ; // column counts
16    cs_cumsum (Cp, w, A->m) ; // column pointers
17    for (j = 0 ; j < A->n ; j++)
18    {
19        for (p = A->p [j] ; p < A->p [j+1] ; p++)
20        {
21            Ci [q = w [Ti [p]]++] = j ; // place A(i,j) as entry C(j,i)
22            if (Cx) Cx [q] = A->x [p] ;
23        }
24    }
25    return (cs_done (T, NULL, w, 1)) ; // success; free w and return
26 }
27
28 // Language: cpp
29 // Compare this snippet from Untitled-1.cpp:
30 //
31 // cs *cs_transpose (const cs *A, int values)
32 // {
33 //     int p, q, j, *Cp, *Ci, *w, *Ti, *Tj ;
34 //     double *Cx, *Tx ;
35 //     cs *T ;
36 //     if (!CS_CSC (A) || !CS_TRIPLET (A)) return (NULL) ; // check inputs
37 //     T = cs_spalloc (A->n, A->m, A->nzmax, values && A->x, 0) ;
38 //     w
39 //
40 =====
```

Is Copilot fair use?

ARTIFICIAL INTELLIGENCE / TECH / LAW

The lawsuit that could rewrite the rules of AI copyright



The key question in the lawsuit is whether open-source code can be reproduced by AI without attached licenses. Credit: Getty Images

/ Microsoft, GitHub, and OpenAI are being sued for allegedly violating copyright law by reproducing open-source code using AI. But the suit could have a huge impact on the wider world of artificial intelligence.

By JAMES VINCENT

Nov 8, 2022, 10:09 AM CST | [9 Comments](#) / [9 New](#)



“AI Art” models: An unfinished story

- Is the training process fair use?
- Can artists refuse to have their art in the training set?
- Do RGPD violations in the LAION-5B dataset appear in trained models?
- No copyright for machine-generated art: how to monetize?
- Can publicly available models trained on LAION-5B be controlled?

Clip Studio Paint will no longer implement an image generator function

December 2, 2022

We apologize for the anxiety and concern we have caused regarding the announcement of the Image Generator Palette on November 29. After the initial announcement, we received a lot of feedback from the community and will no longer implement the image generator palette.

Here are some concerns that we have taken to heart:

- Current image generation AI exploits other artists' intellectual property and is unusable
- This feature will hinder rather than help artists in their creative activities
- Using artist's work that is not opted-in to a data set is ethically unacceptable
- The fear that this will make Clip Studio Paint artwork synonymous with AI-generated work
- There are existing features that need to be prioritized over image generation AI features
- Having something unknown in the app I use daily potentially infringing on legal or moral rights is unacceptable
- Clip Studio Paint should be an app that takes responsibility for a safe, creative environment

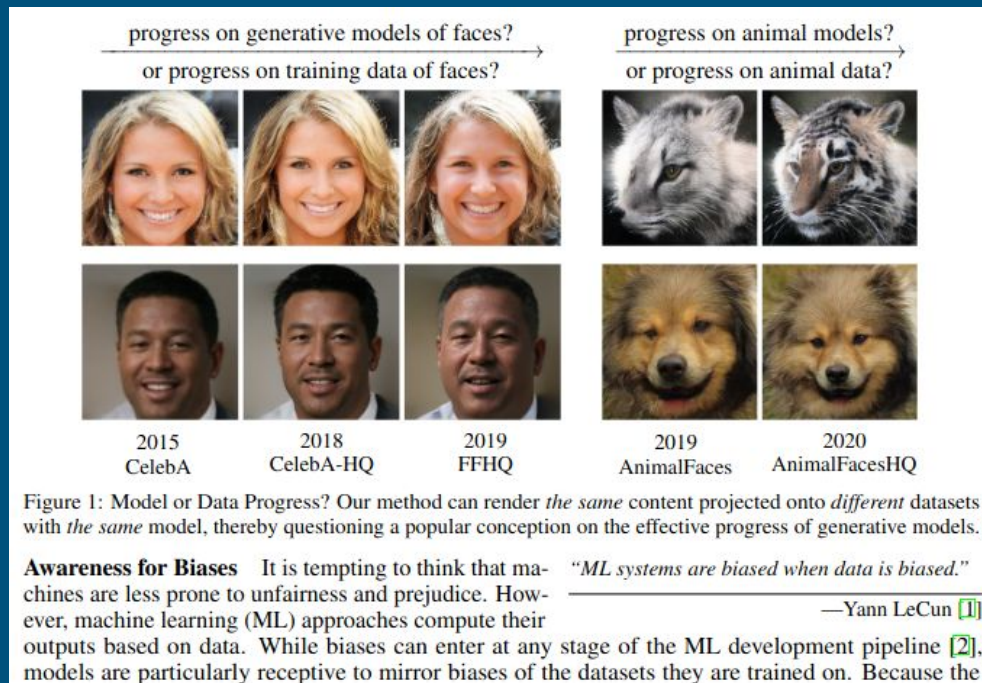
The Image Generator palette was developed with the hope that we could provide a new way to experience creativity. However, we had neglected to take the necessary considerations.

We were so preoccupied with how generative AI technology could be used creatively that we lost sight of what our core users want from Clip Studio Paint as a creative tool. We would like to sincerely apologize.

Celsys will not be implementing any image generator AI features in Clip Studio Paint that present this kind of concern, and will listen more sincerely to the opinions of creators. We will redouble our efforts to provide features that can be used with peace of mind by all artists.

Existing problems with image generation

- Falsification of evidence
 - Ease of access to fake images
 - Revenge pornography
 - Data bias
-
- Problems identified since ~2014, few solutions



Esser, Patrick, Robin Rombach, and Björn Ommer. "A note on data biases in generative models." arXiv preprint arXiv:2012.02516 (2020).

Deep generative models are for more than images

- Rüttgers, Mario, et al. "**Prediction of a typhoon track** using a generative adversarial network and satellite images." Scientific reports 9.1 (2019): 1-15.
- Salgueiro Romero, Luis, Javier Marcello, and Verónica Vilaplana. "**Super-resolution of sentinel-2 imagery** using generative adversarial networks." Remote Sensing 12.15 (2020): 2424.
- Ren, Hailin, and Pinhas Ben-Tzvi. "Learning inverse kinematics and **dynamics of a robotic manipulator** using generative adversarial networks." Robotics and Autonomous Systems 124 (2020): 103386.
- Ingraham, John, et al. "Generative models for graph-based **protein design**." Advances in neural information processing systems 32 (2019).