NATIONAL CENTRE
FOR NUCLEAR RESEARCH
ŚWIERK

# Biased AI image generations models and misconceptions about health conditions

Marianna Zadrożna, Adam Zadrożny

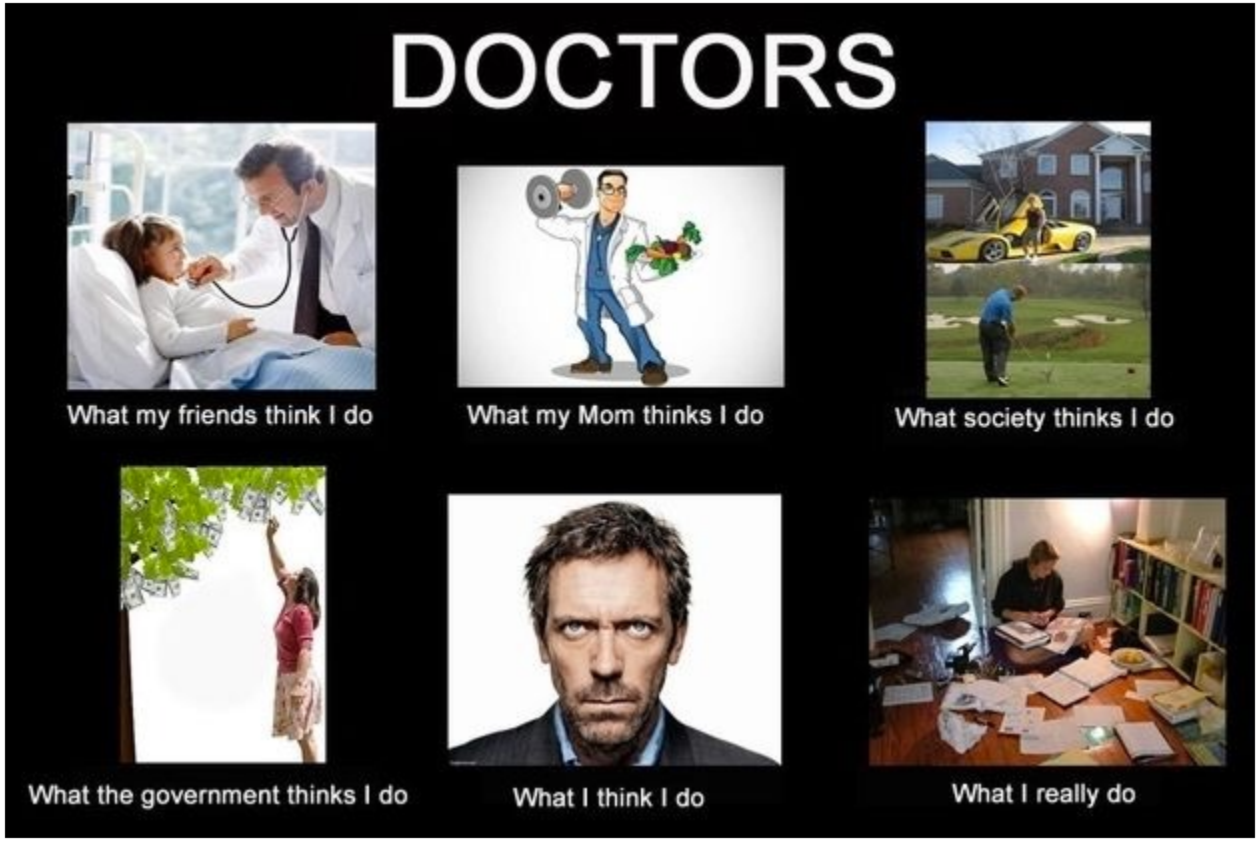
The International Workshop on Machine Learning and Quantum Computing Applications in Medicine and Physics

13-16 September 2022, Warsaw

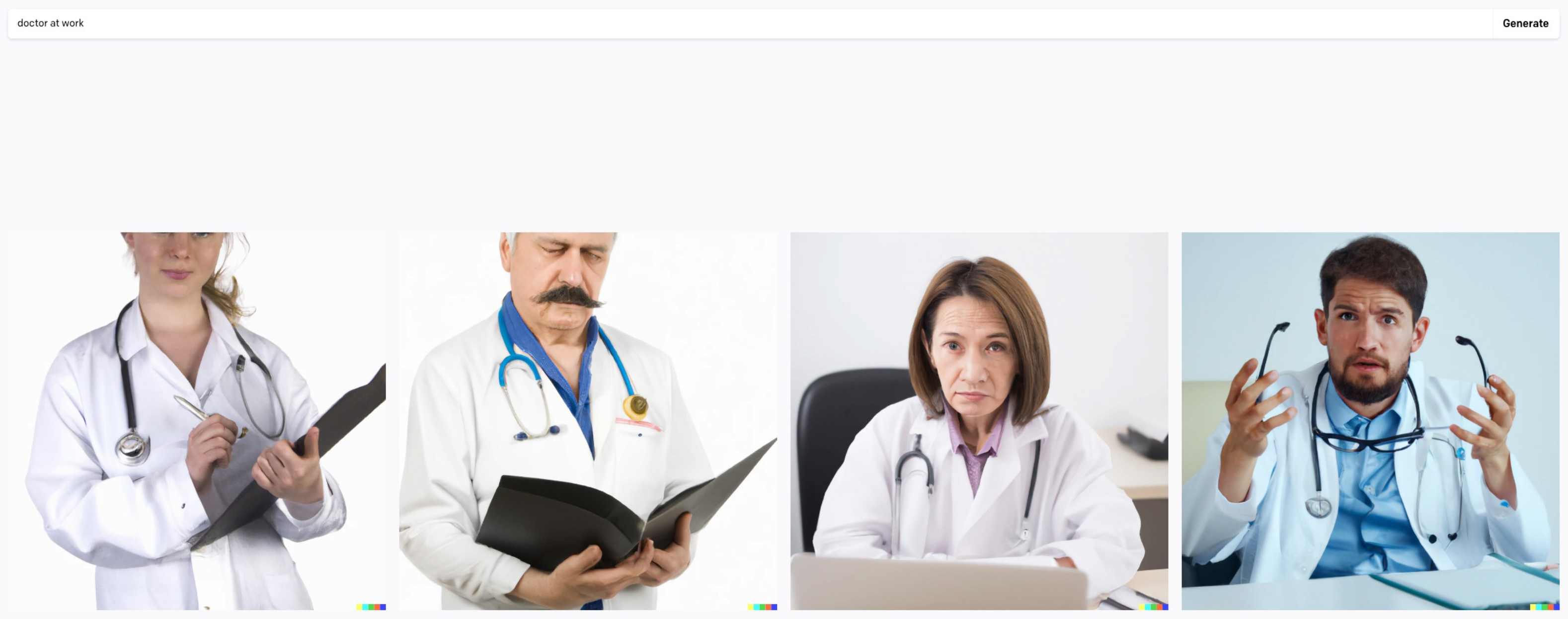
# AI is far from being ideal

People with no idea about AI, telling me my AI will destroy the world
Me wondering why my neural network is classifying a cat as a dog..
Dog

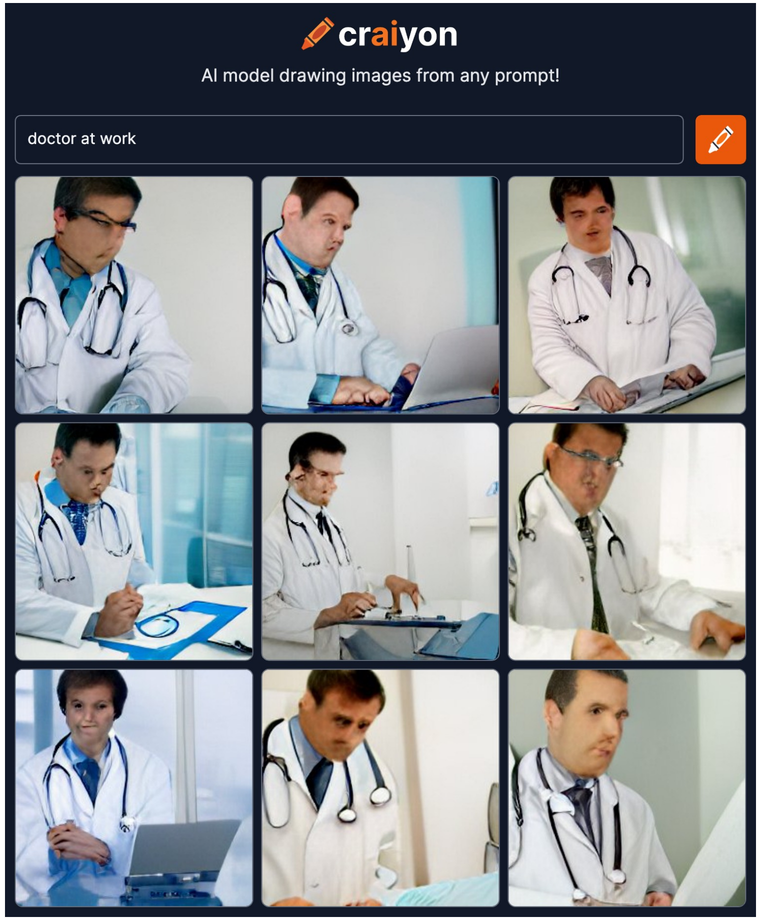
# Our colective imagination or what society things that medics do?



# What AI thinks that medics do?


doctor at work

DALL-E 2

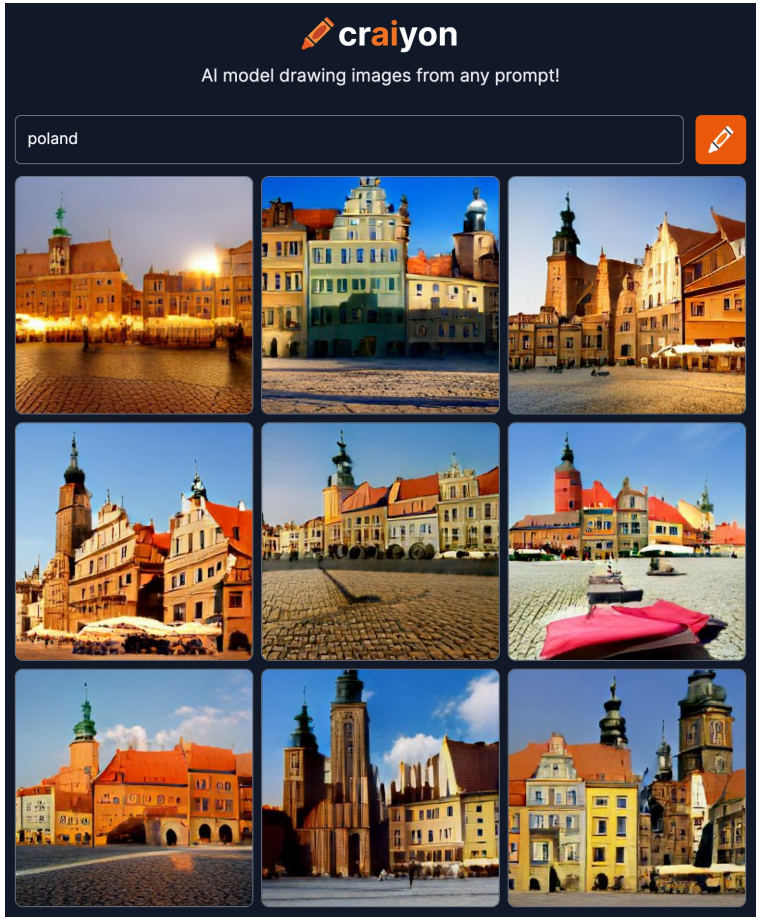


doctor at work

# What medics really do?

- In Poland: Filling up documentation for Narodowy Fundusz Zdrowia (NFZ)

# Raise of Text-to-image models
# Siri! Show me Poland!



# Text-to-image models history

- 2017 – transformers model - Google
  - arXiv:1706.03762
- 2018 – Generative Pre-trained Transformer (GPT) – OpenAI
  - https://www.gwern.net/docs/www/s3-us-west-2.amazonaws.com/d73fdc5ffa8627bce44dcda2fc012da638ffb158.pdf
- 2020 – Generative Pre-trained Transformer 3 (GPT 3)
  - arXiv:2005.14165
- 2020 – DALL-E - OpenAI
  - arXiv:2102.12092
- 2022 – DALL-E 2 - OpenAI
  - https://cdn.openai.com/papers/dall-e-2.pdf
- 2022 – DALL-E mini, Midjourney, Stable Diffusion

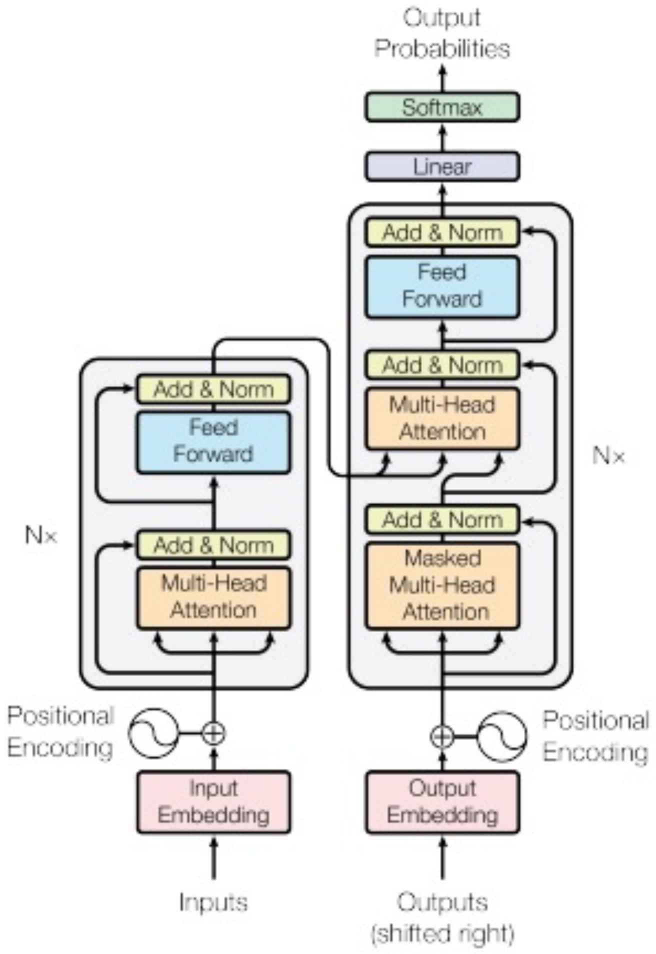
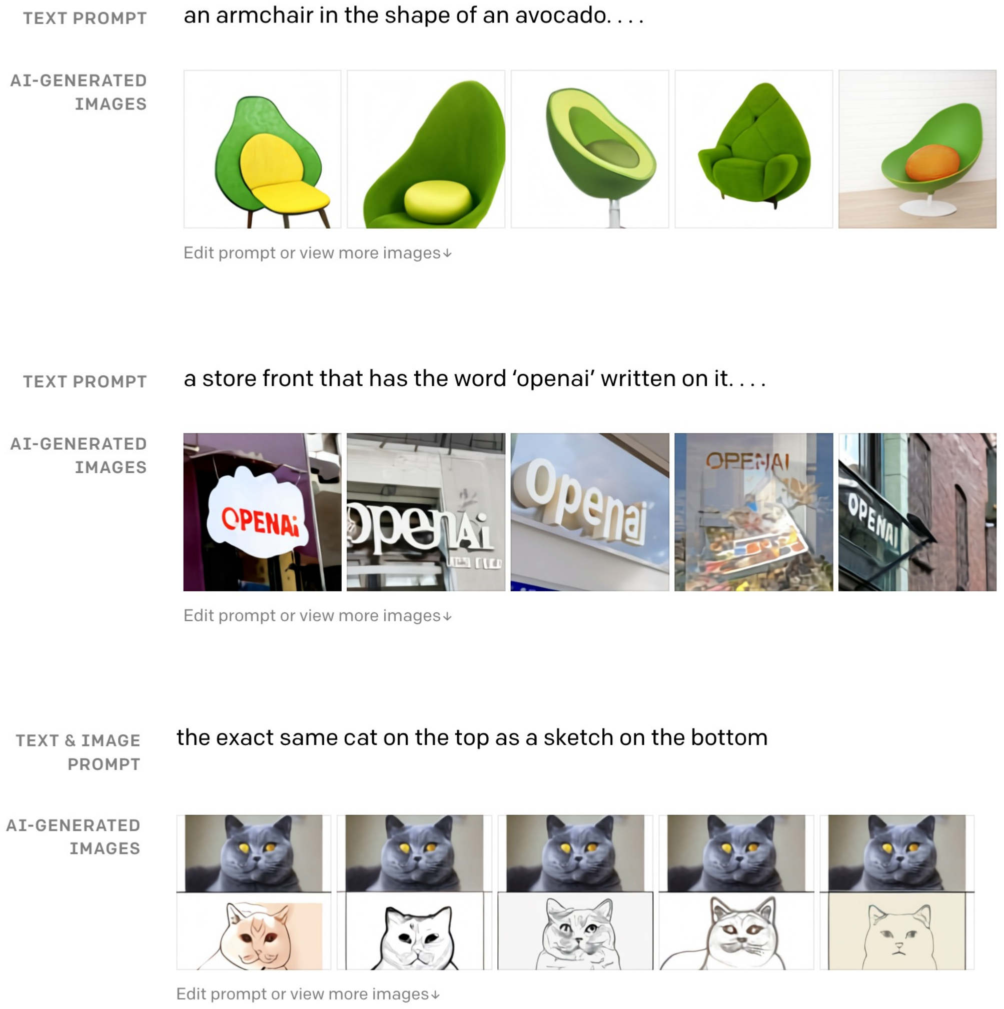


Figure 1: The Transformer - model architecture.

Source: arXiv:1706.03762

# DALL-E

- DALL-E comes from joining name Salvator Dali and WALL-E
- Trained on 400 M pairs of images and descriptions
- https://openai.com/blog/dall-e/
- https://arxiv.org/abs/2102.12092



TEXT PROMPT an armchair in the shape of an avocado. . . .

AI-GENERATED IMAGES

TEXT PROMPT a store front that has the word 'openai' written on it. . . .

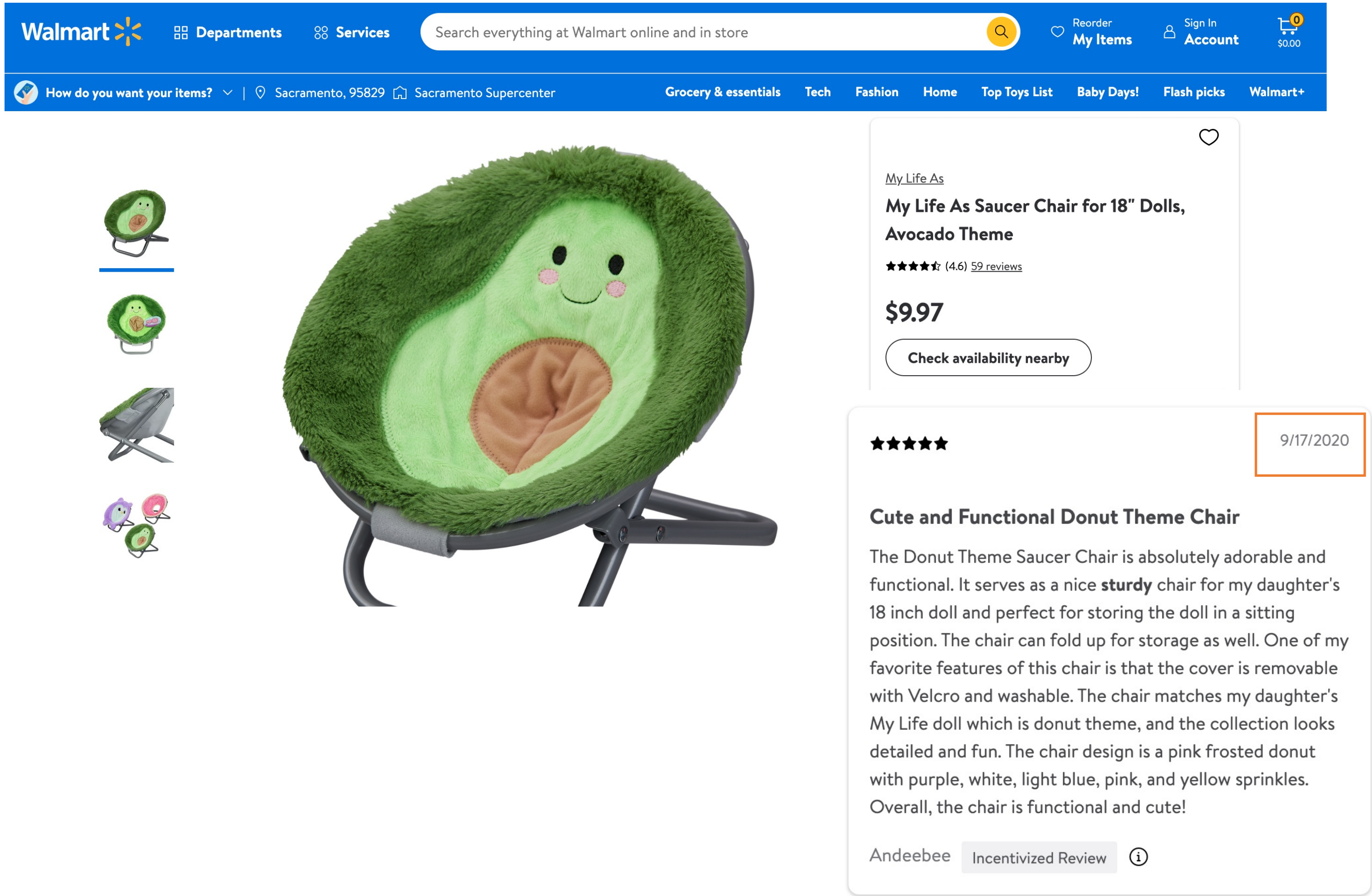
AI-GENERATED IMAGES

TEXT & IMAGE PROMPT the exact same cat on the top as a sketch on the bottom

AI-GENERATED IMAGES

Edit prompt or view more images↓
OPENAi
openAi
Openai
OPENAI
OPENAI
Edit prompt or view more images↓
Edit prompt or view more images↓

# Avocado Armchair exists!



My Life As

My Life As Saucer Chair for 18" Dolls, Avocado Theme

★★★★☆ (4.6) 59 reviews

$9.97

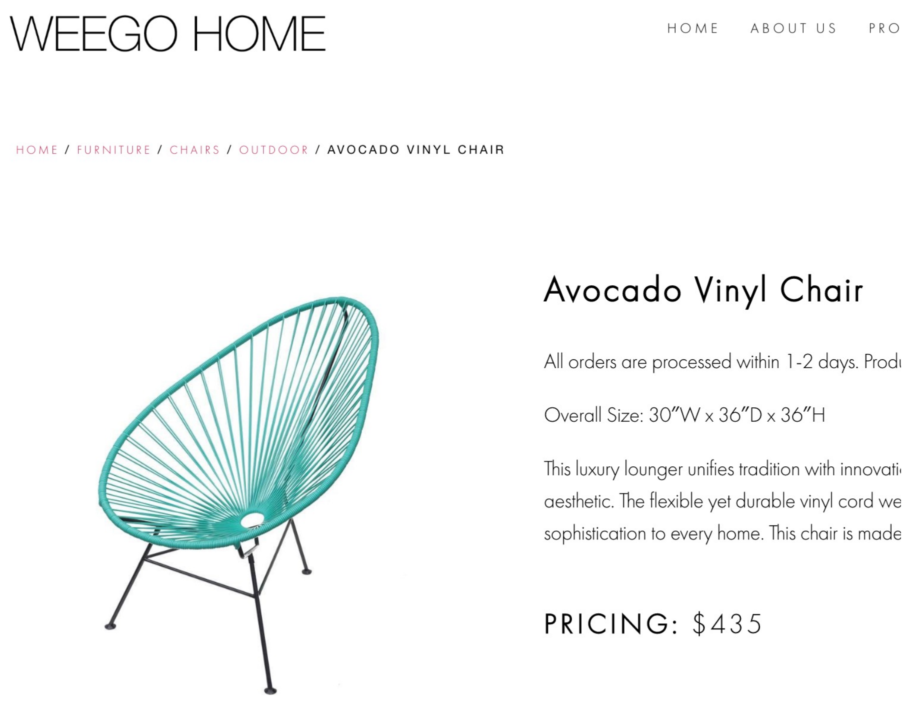
Check availability nearby

★★★★★ 9/17/2020

Cute and Functional Donut Theme Chair

The Donut Theme Saucer Chair is absolutely adorable and functional. It serves as a nice **sturdy** chair for my daughter's 18 inch doll and perfect for storing the doll in a sitting position. The chair can fold up for storage as well. One of my favorite features of this chair is that the cover is removable with Velcro and washable. The chair matches my daughter's My Life doll which is donut theme, and the collection looks detailed and fun. The chair design is a pink frosted donut with purple, white, light blue, pink, and yellow sprinkles. Overall, the chair is functional and cute!

Andeebee Incentivized Review ⓘ



HOME / FURNITURE / CHAIRS / OUTDOOR / AVOCADO VINYL CHAIR

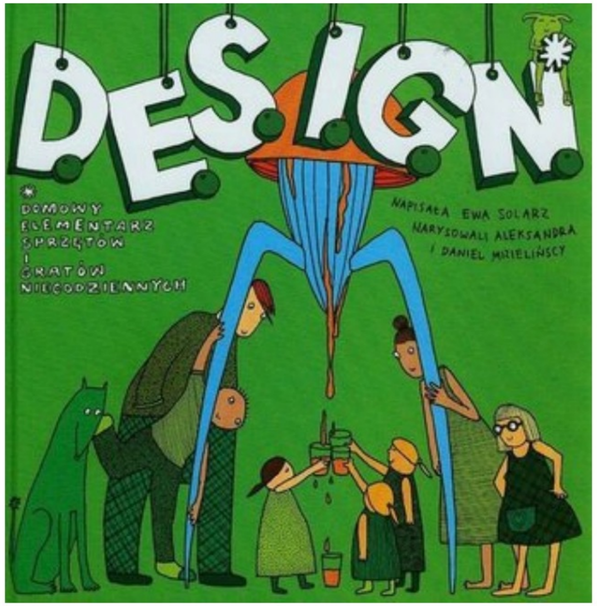
Avocado Vinyl Chair

All orders are processed within 1-2 days. Prod

Overall Size: 30″W x 36″D x 36″H

This luxury lounger unifies tradition with innovati aesthetic. The flexible yet durable vinyl cord we sophistication to every home. This chair is made

PRICING: $435



DESIGN
DOMOWY ELEMENTARZ SPRZĘTÓW I GRATÓW NIECODZIENNYCH
NAPISAŁA EWA SOLARZ
NARYSOWALI ALEKSANDRA I DANIEL MIZIELIŃSCY

# DALL-E 2

- Published in 2022 by OpenAI
- CLIP - Contrastive Language-Image Pre-training
- https://cdn.openai.com/papers/dall-e-2.pdf

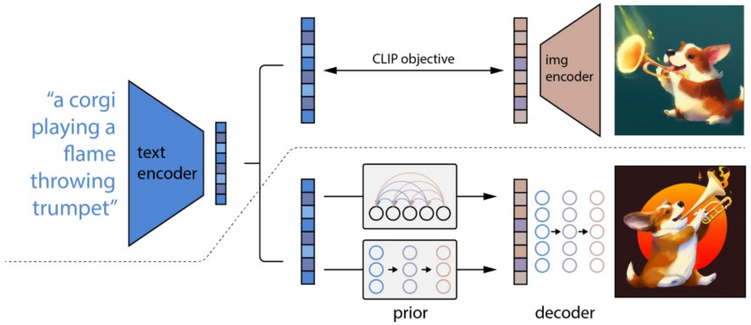


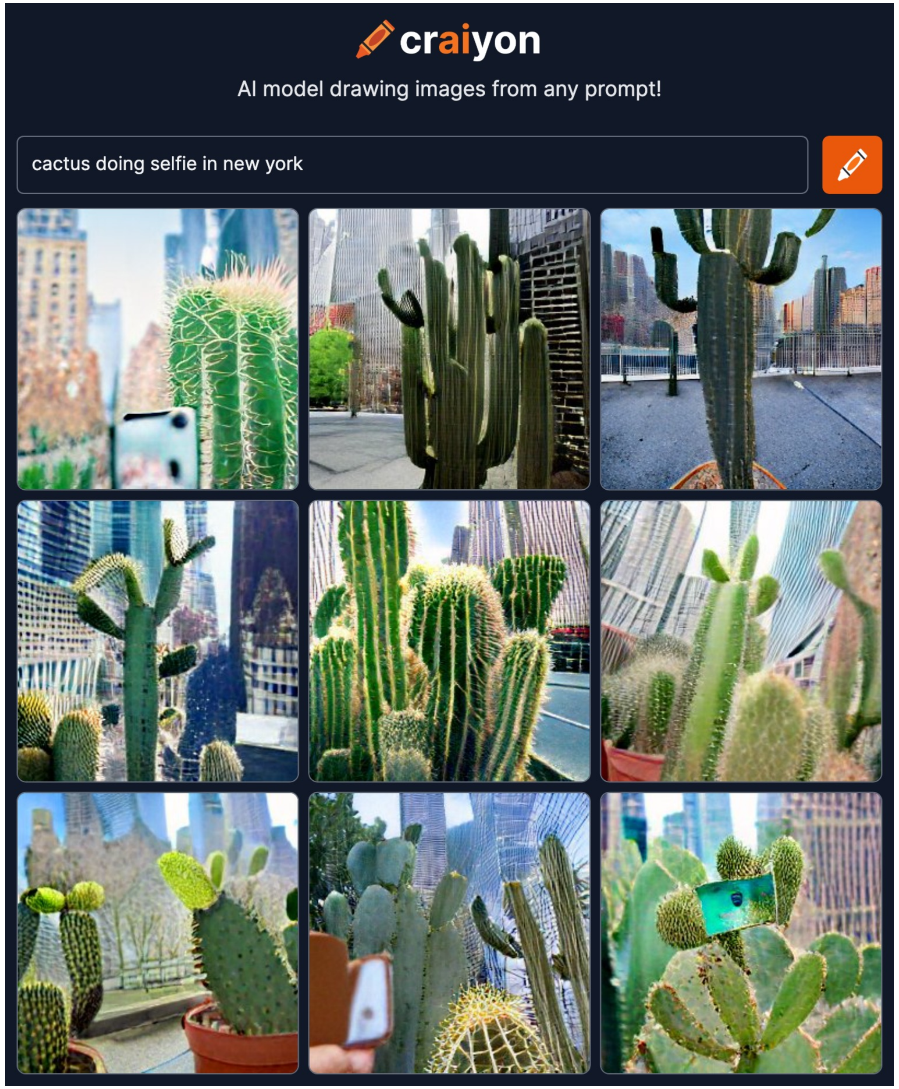
Figure 2: A high-level overview of unCLIP. Above the dotted line, we depict the CLIP training process, through which we learn a joint representation space for text and images. Below the dotted line, we depict our text-to-image generation process: a CLIP text embedding is first fed to an autoregressive or diffusion prior to produce an image embedding, and then this embedding is used to condition a diffusion decoder which produces a final image. Note that the CLIP model is frozen during training of the prior and decoder.

DALL-E 2

Teddy bears working on new AI research underwater with 1990s technology

# DALLE-mini a.k.a. craiyon.com

- Created by Craiyon LLC
- Trained on Google TRC
- Source of many memes
- Paints faces really badly, you will see later ...

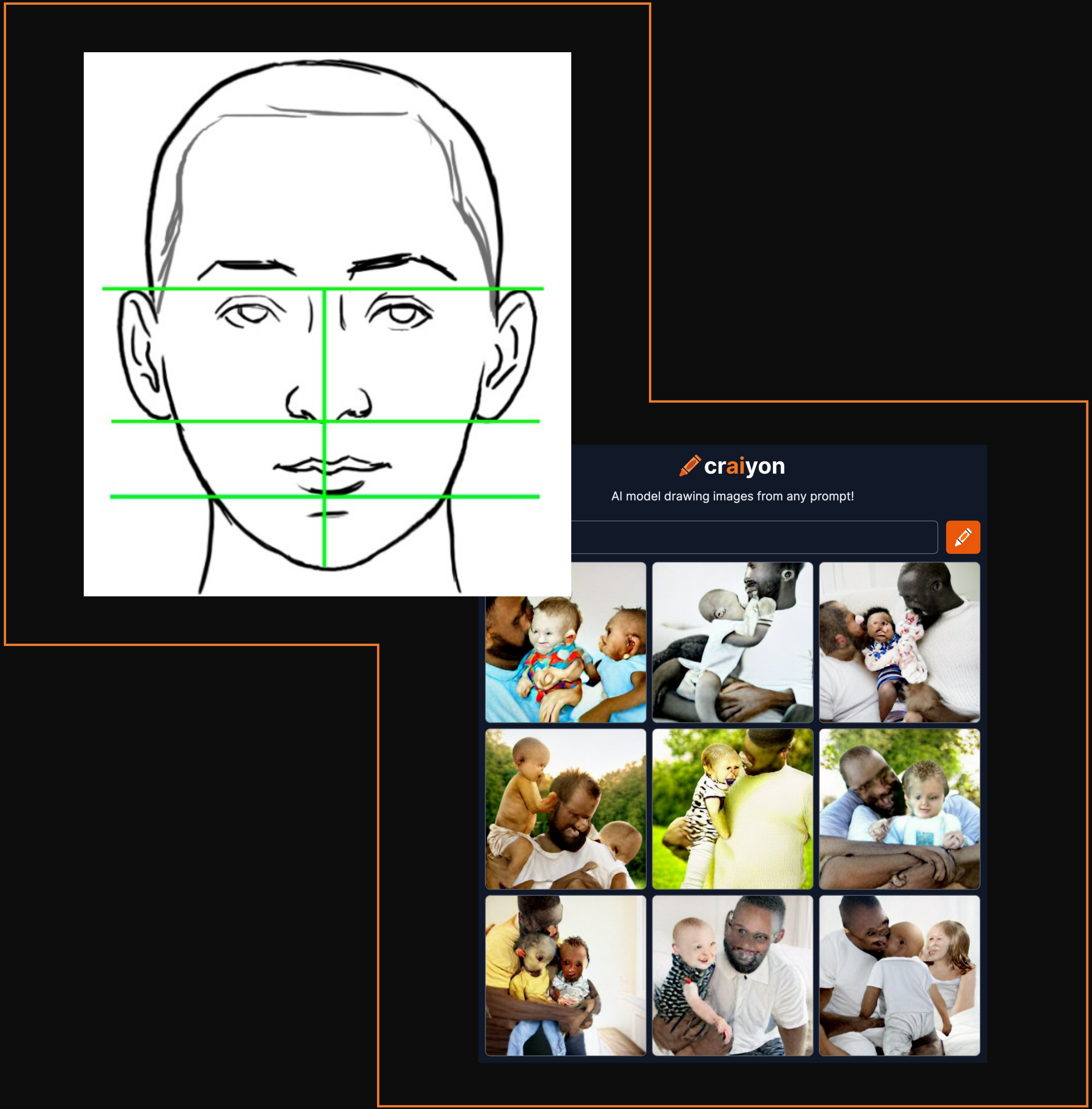


cactus doing selfie in new york

# Stuff AI has to learn is to draw a proper face

It is really hard to explain to AI some algorithm concepts like drawing a face.

The best is stable diffusion or some GAN networks dedicated to faces.



# Midjourney

- Created by company Midjourney

https://www.howtogeek.com/823337/how-to-create-synthetic-ai-art-with-midjourney/

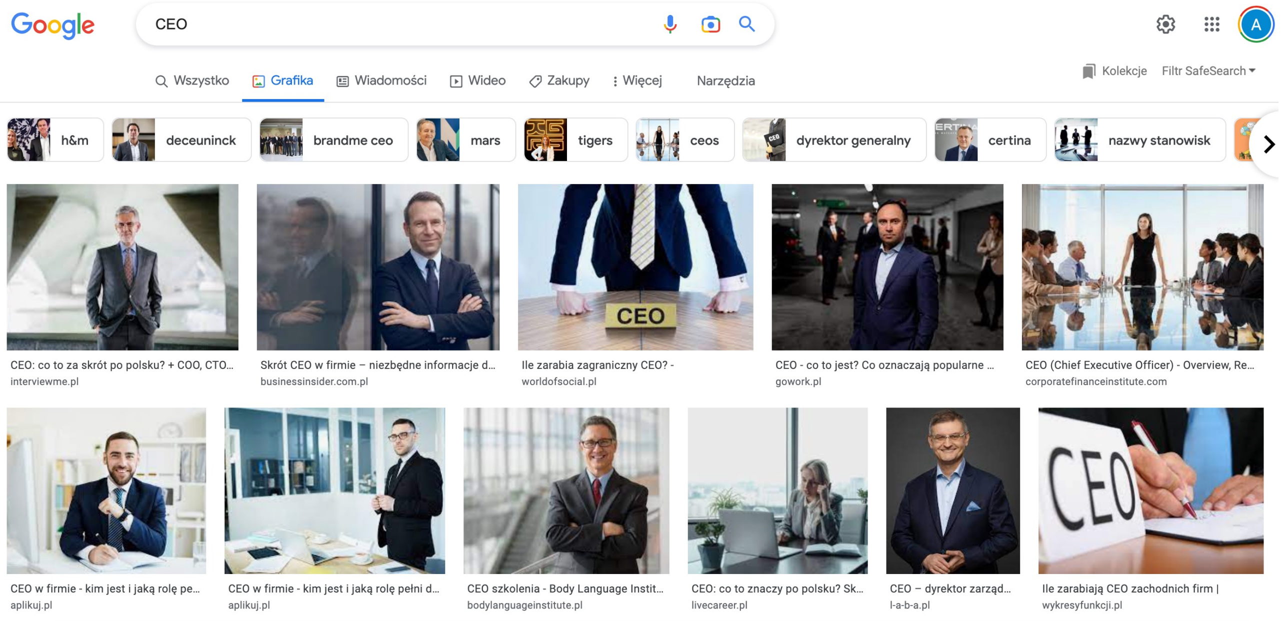
Mechanical dove

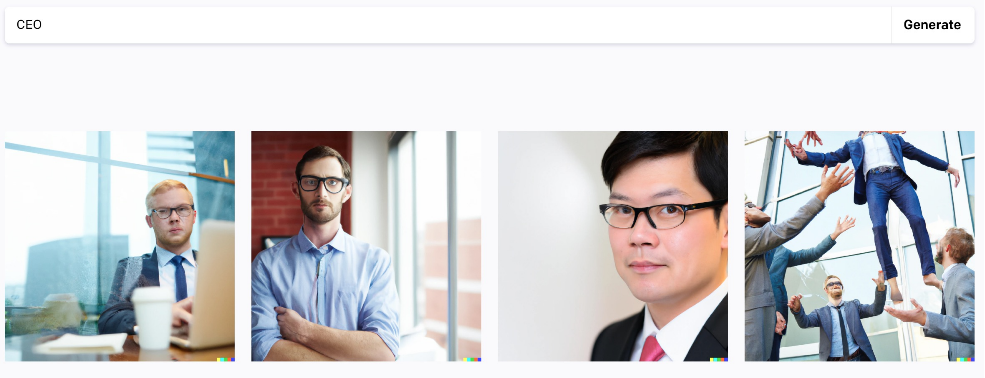
# Stable Diffusion

- Developed by StaticAI
  - Collaboration between EleutherAI and LAION
- Finally faces look realistic
- https://github.com/CompVis/stable-diffusion

# Bias and debiasing of AI models

- We live in biased world
- Media and blog post are biased one way or the other
- To train the models we need a lot of data
- In order to the bias we need to have a balanced ratio of gender, race, age, …
- And we can force it on the output

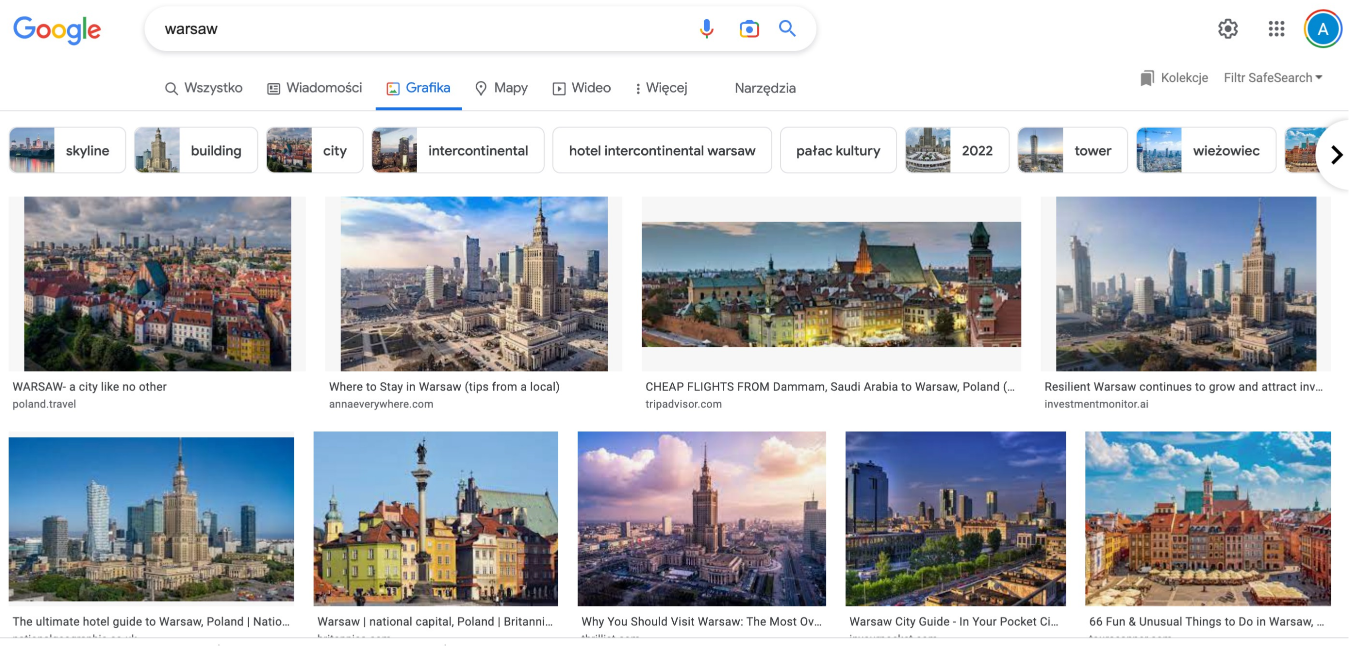

Google
CEO
Wszystko
Grafika
Wiadomości
Wideo
Zakupy
Więcej
Narzędzia
Kolekcje
Filtr SafeSearch
h&m
deceuninck
brandme ceo
mars
tigers
ceos
dyrektor generalny
certina
nazwy stanowisk
CEO: co to za skrót po polsku? + COO, CTO...
interviewme.pl
Skrót CEO w firmie – niezbędne informacje d...
businessinsider.com.pl
Ile zarabia zagraniczny CEO? -
worldofsocial.pl
CEO - co to jest? Co oznaczają popularne ...
gowork.pl
CEO (Chief Executive Officer) - Overview, Re...
corporatefinanceinstitute.com
CEO w firmie - kim jest i jaką rolę pe...
aplikuj.pl
CEO w firmie - kim jest i jaką rolę pełni d...
aplikuj.pl
CEO szkolenia - Body Language Instit...
bodylanguageinstitute.pl
CEO: co to znaczy po polsku? Sk...
livecareer.pl
CEO – dyrektor zarząd...
l-a-b-a.pl
Ile zarabiają CEO zachodnich firm |
wykresyfunkcji.pl



DALL-E 2

# Why we have bias in AI?

- Traning data



WARSAW- a city like no other
poland.travel

Where to Stay in Warsaw (tips from a local)
annaeverywhere.com

CHEAP FLIGHTS FROM Dammam, Saudi Arabia to Warsaw, Poland (...
tripadvisor.com

Resilient Warsaw continues to grow and attract inv...
investmentmonitor.ai

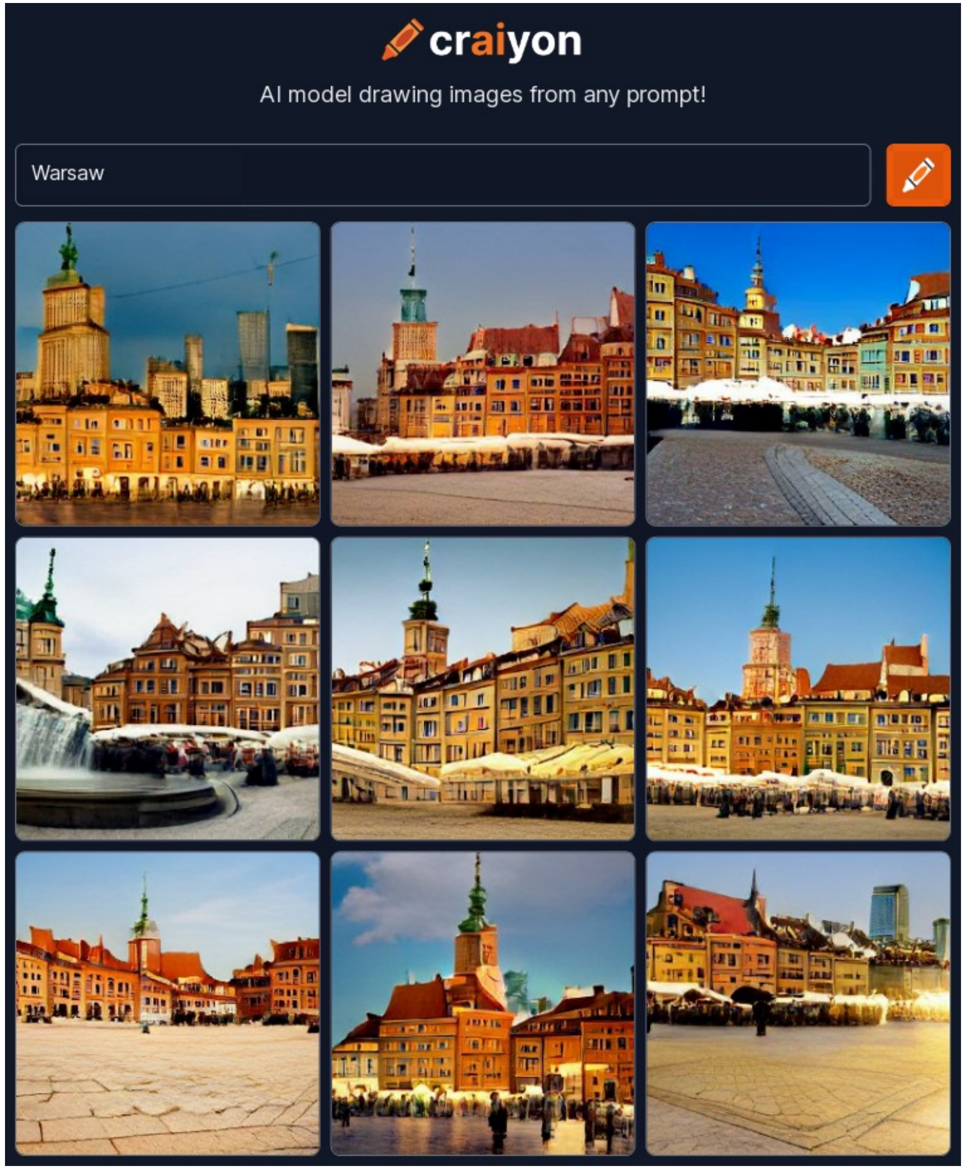
The ultimate hotel guide to Warsaw, Poland | Natio...

Warsaw | national capital, Poland | Britanni...

Why You Should Visit Warsaw: The Most Ov...

Warsaw City Guide - In Your Pocket Ci...

66 Fun & Unusual Things to Do in Warsaw, ...



AI model drawing images from any prompt!

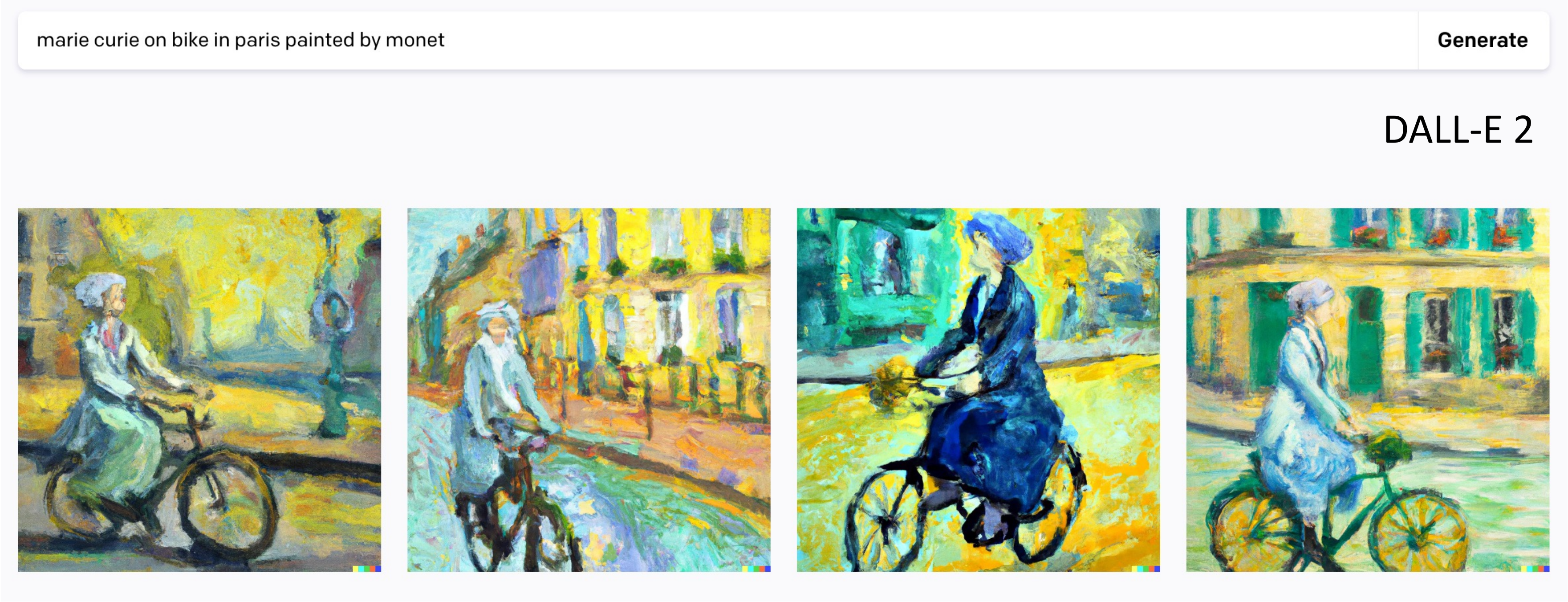
Warsaw

# Main research hypothesis:

By asking AI to fill the blanks we can probe biases in training data and this leads to probing biases in media space

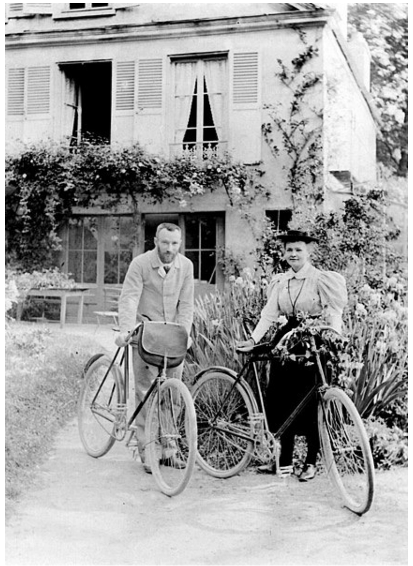
# Using AI to study misconceptions, biases and hidden assumtions in infosphere

- If AI can generate image from text in has to have some connection between words and real world objects
- If we are not precise that elements that we did not specify has be filled from intuition derived from dataset



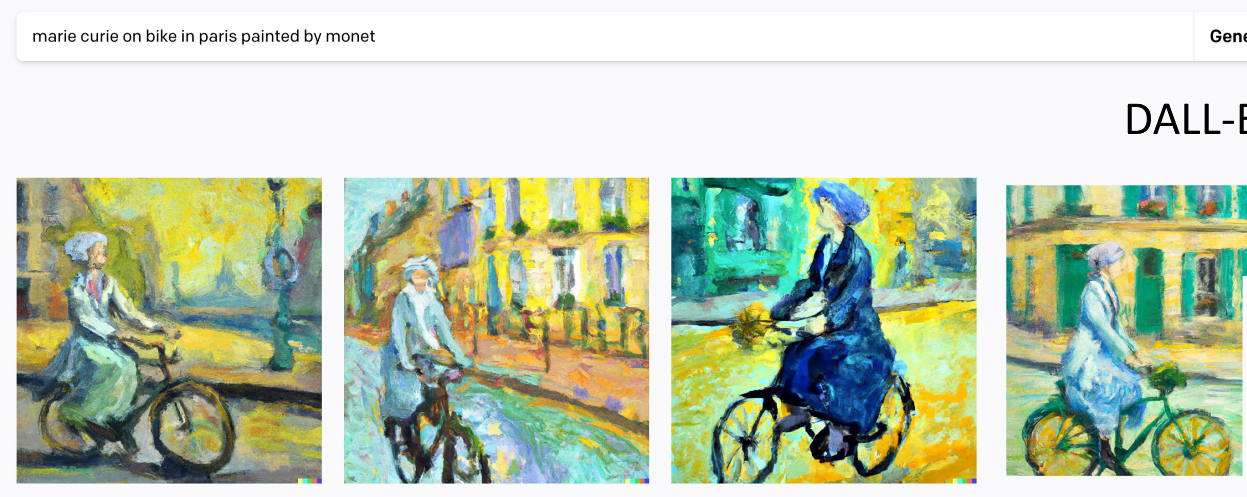
marie curie on bike in paris painted by monet

DALL-E 2



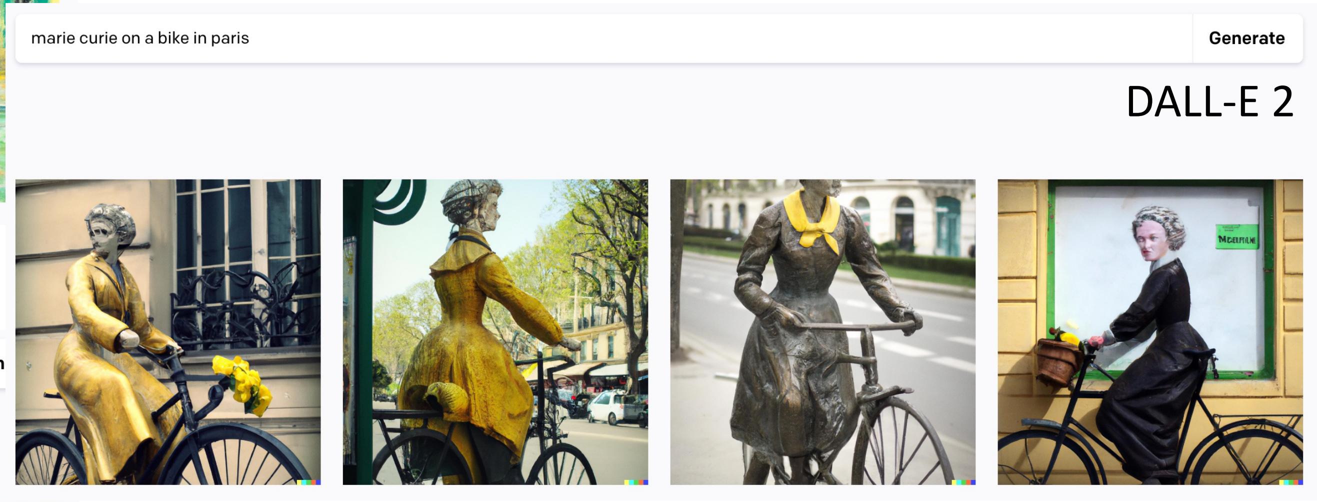
https://commons.wikimedia.org/wiki/File:Pierre_et_Marie_Curie_devant_leur_maison_de_Sceaux_en_1895.jpg

# Using AI to study misconceptions, biases and hidden assumtions in infosphere



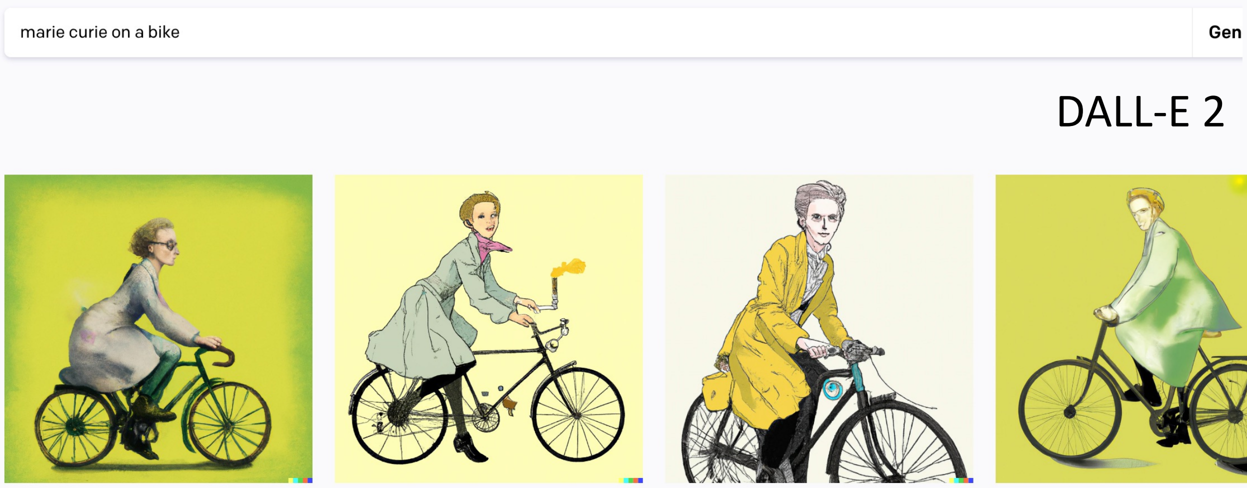
marie curie on bike in paris painted by monet

DALL-E 2



marie curie on a bike in paris

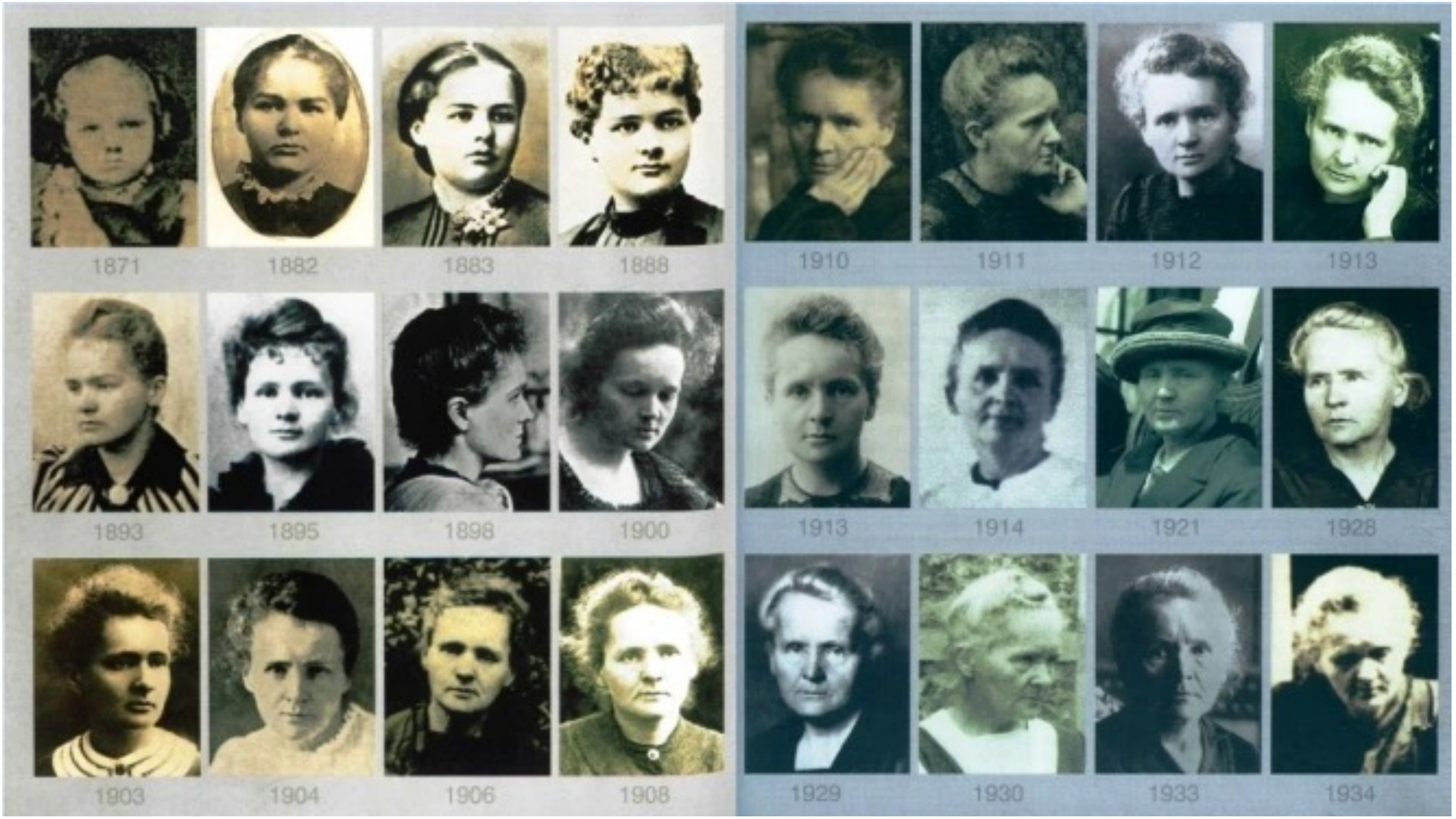
DALL-E 2



marie curie on a bike

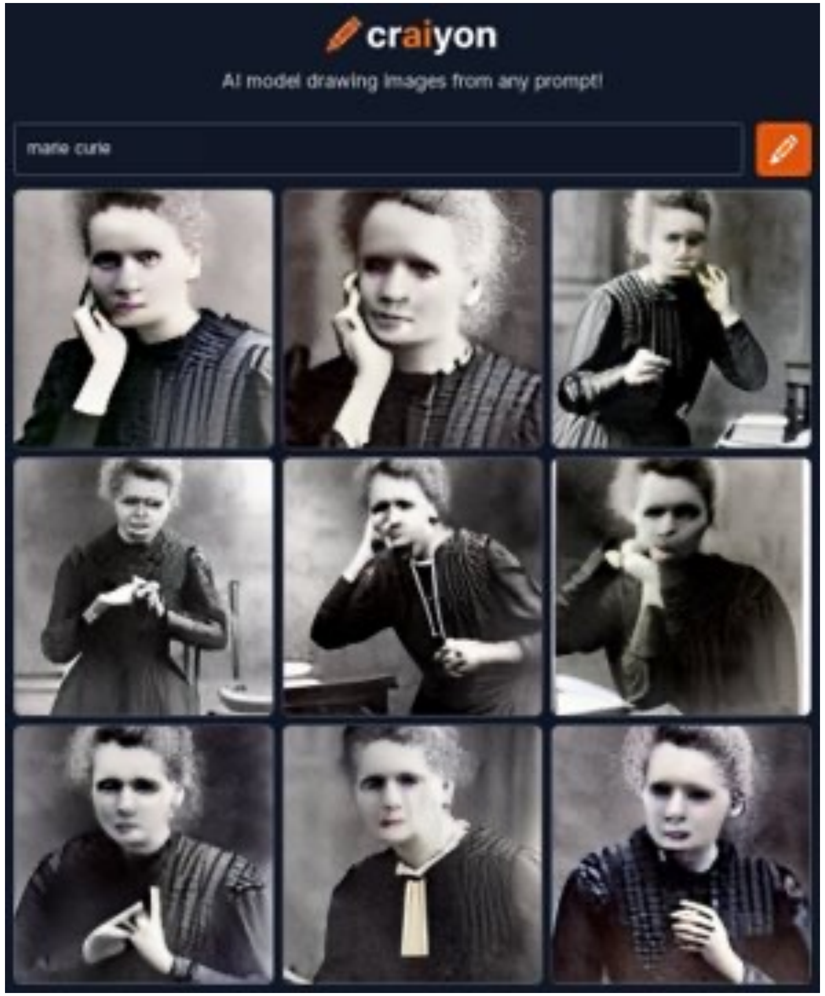
DALL-E 2

# More funny stuff …



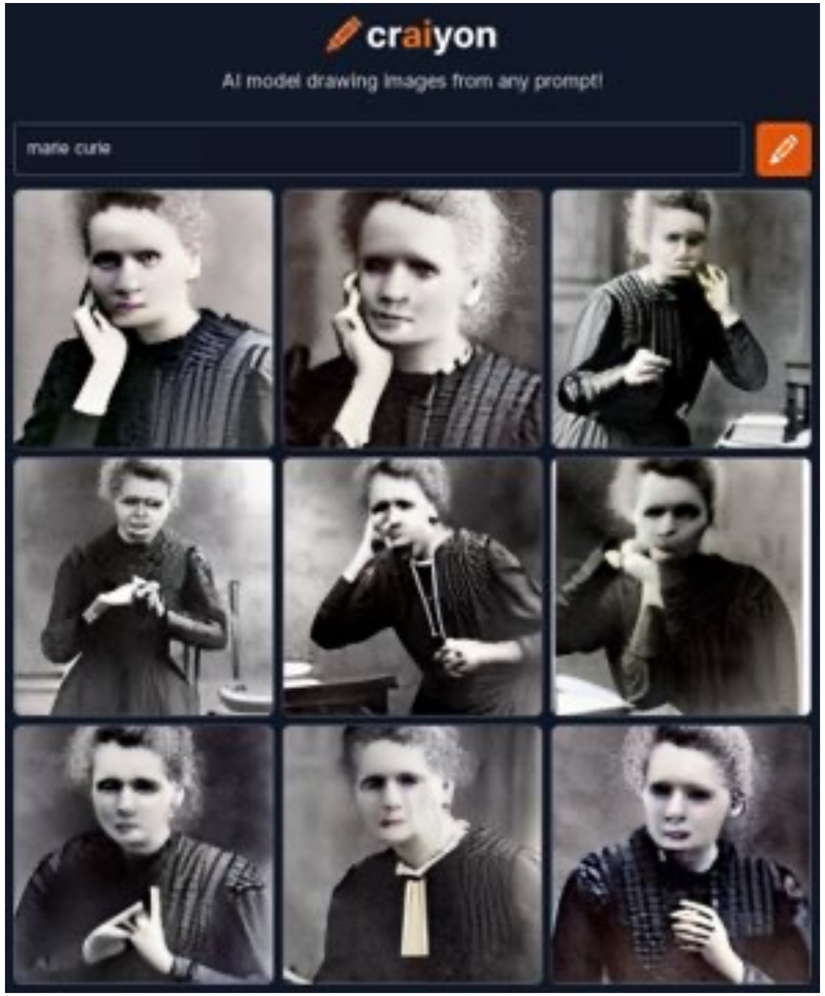
Maria Curie-Skłodowska, over the years
https://niezbednikchemika.files.wordpress.com/2017/10/miniatury-msc.jpg

# More funny stuff …

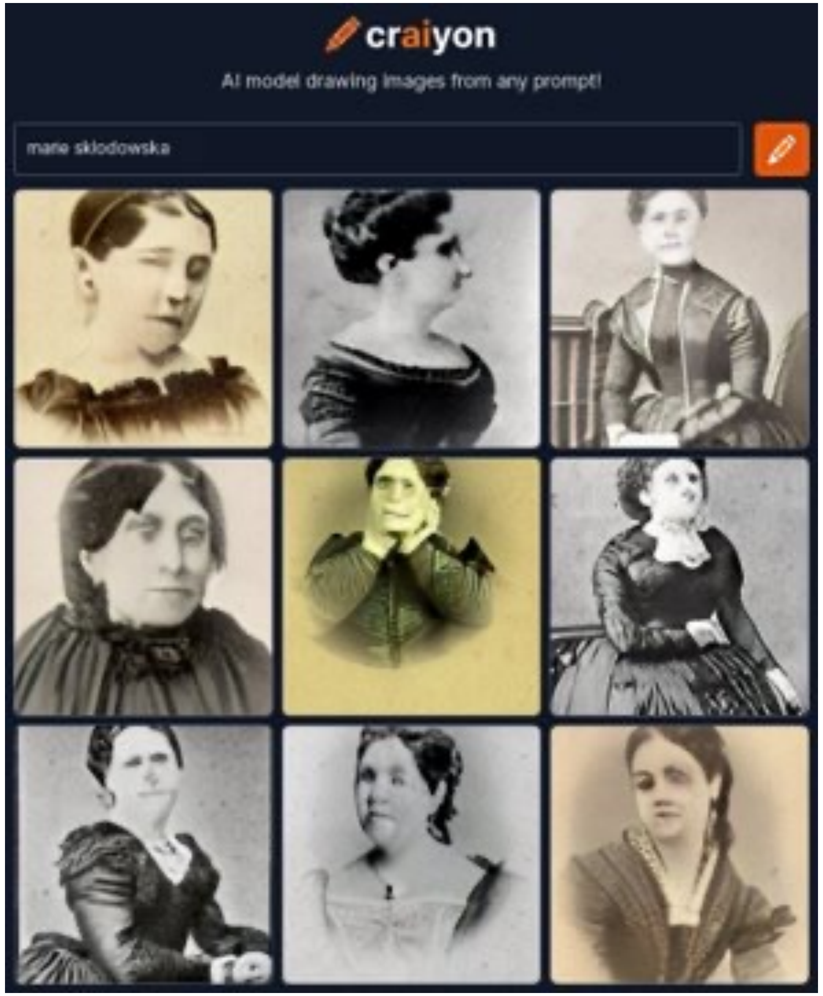


marie curie

# More funny stuff …

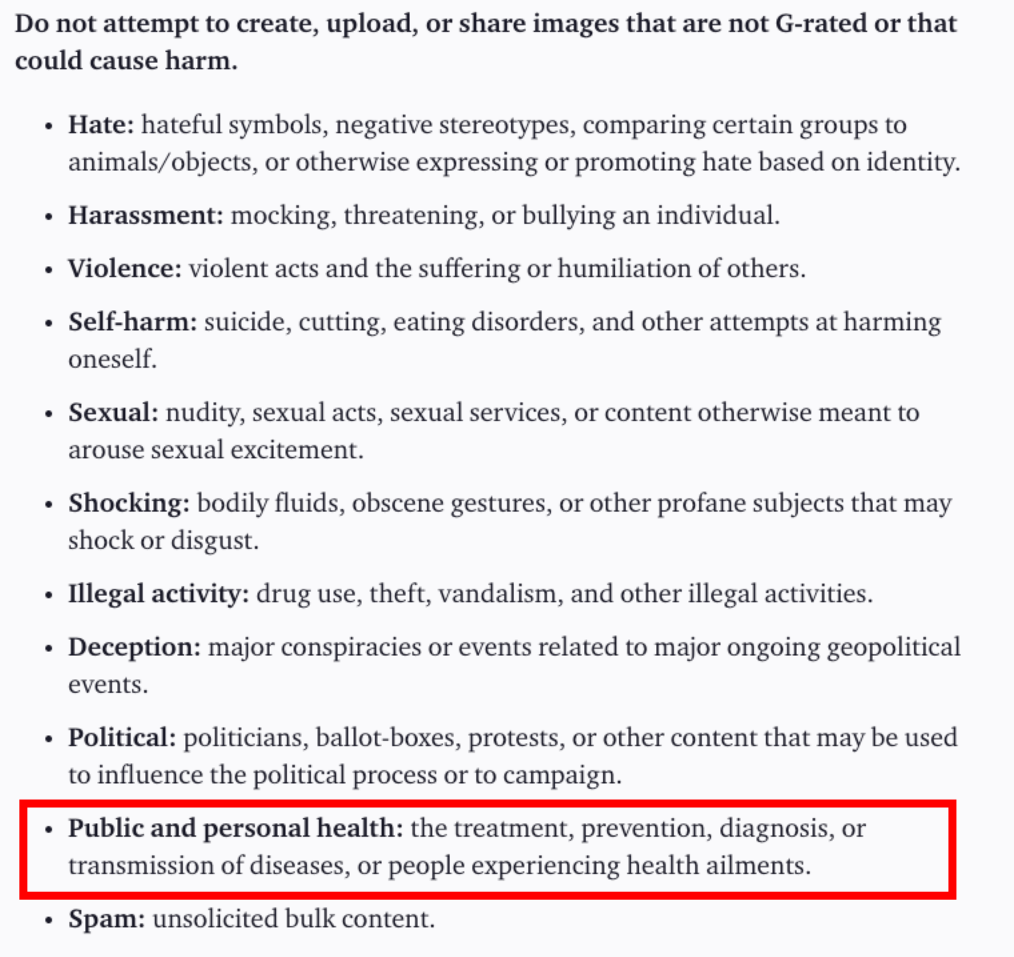


marie curie



marie sklodowska

# Limitations of DALL-E 2

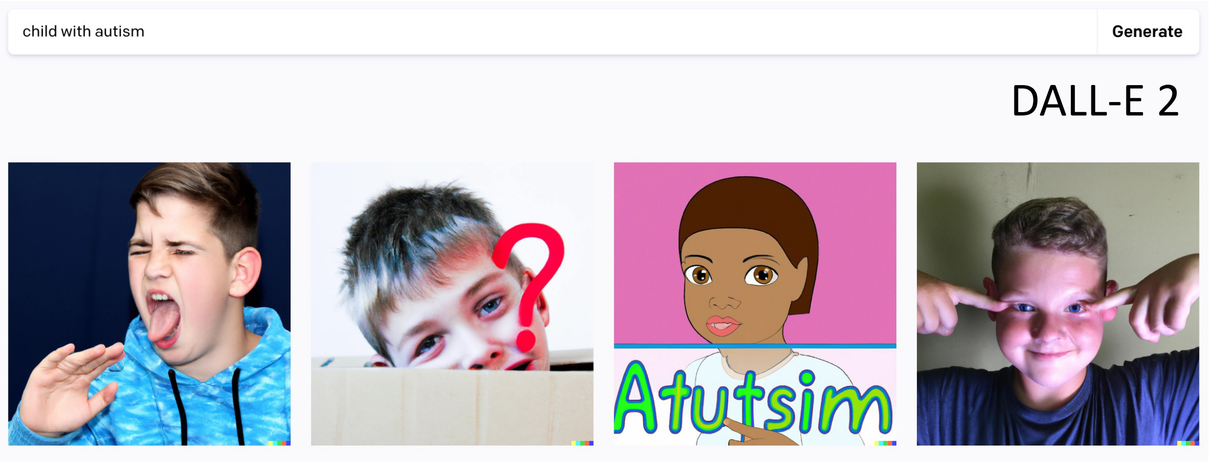


**Do not attempt to create, upload, or share images that are not G-rated or that could cause harm.**

- **Hate:** hateful symbols, negative stereotypes, comparing certain groups to animals/objects, or otherwise expressing or promoting hate based on identity.
- **Harassment:** mocking, threatening, or bullying an individual.
- **Violence:** violent acts and the suffering or humiliation of others.
- **Self-harm:** suicide, cutting, eating disorders, and other attempts at harming oneself.
- **Sexual:** nudity, sexual acts, sexual services, or content otherwise meant to arouse sexual excitement.
- **Shocking:** bodily fluids, obscene gestures, or other profane subjects that may shock or disgust.
- **Illegal activity:** drug use, theft, vandalism, and other illegal activities.
- **Deception:** major conspiracies or events related to major ongoing geopolitical events.
- **Political:** politicians, ballot-boxes, protests, or other content that may be used to influence the political process or to campaign.
- **Public and personal health:** the treatment, prevention, diagnosis, or transmission of diseases, or people experiencing health ailments.
- **Spam:** unsolicited bulk content.

# Study case: autism

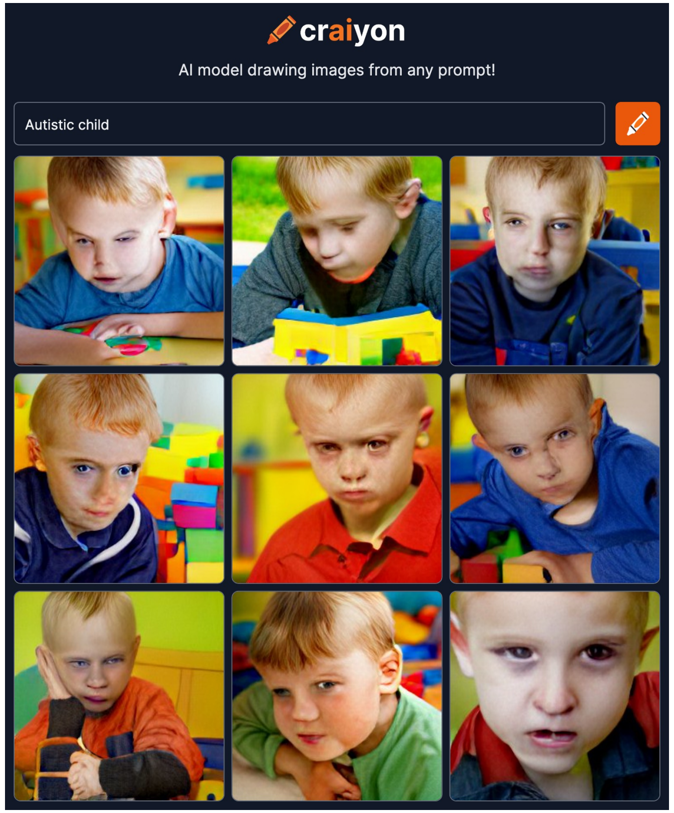
- There was a wide spread assumption that autism is something connected to boys



child with autism

Generate

DALL-E 2

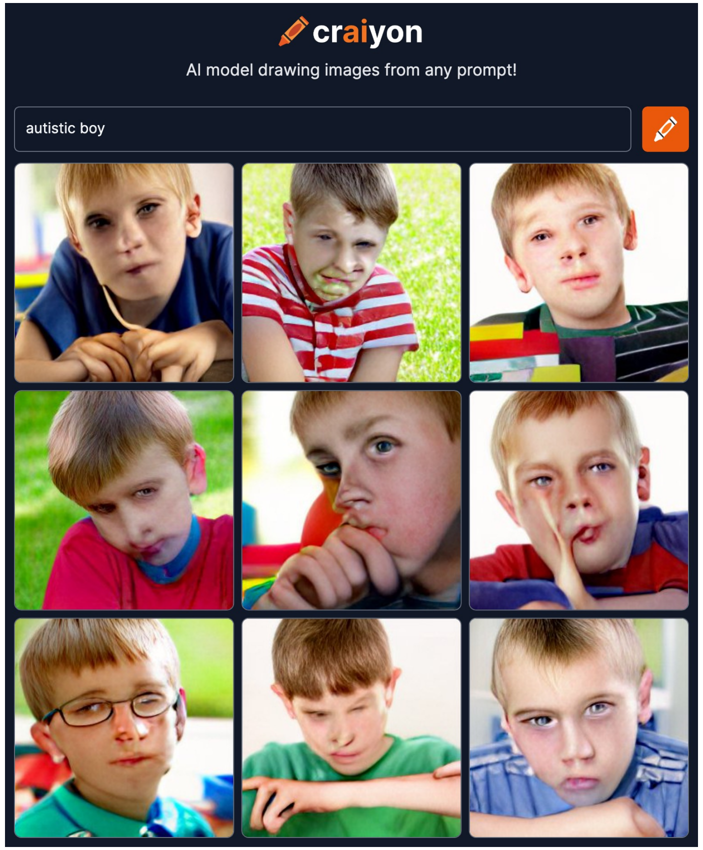


AI model drawing images from any prompt!

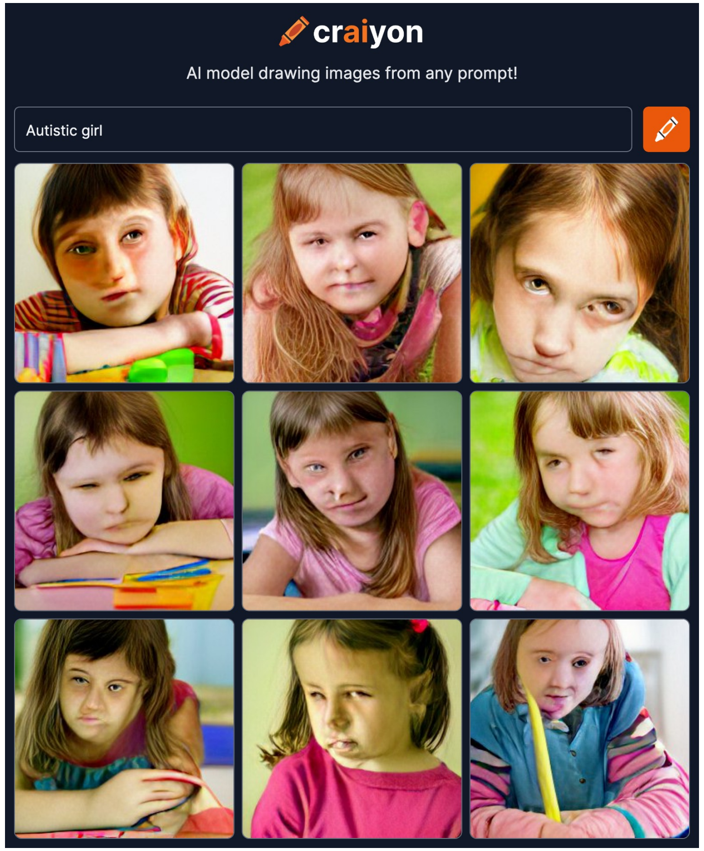
Autistic child

# Study case: autism

- But we can explictly ask for autistic boy or a girl

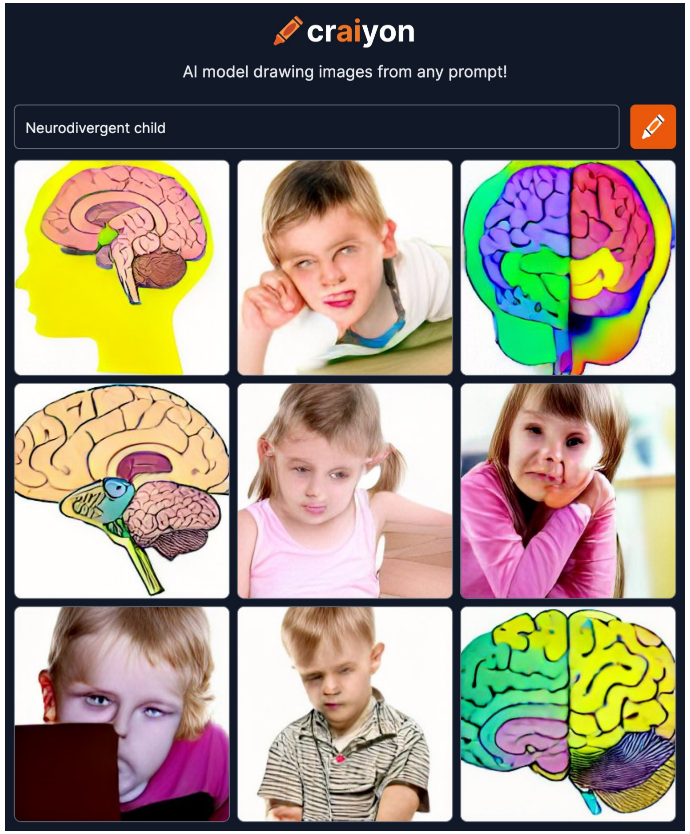


autistic boy

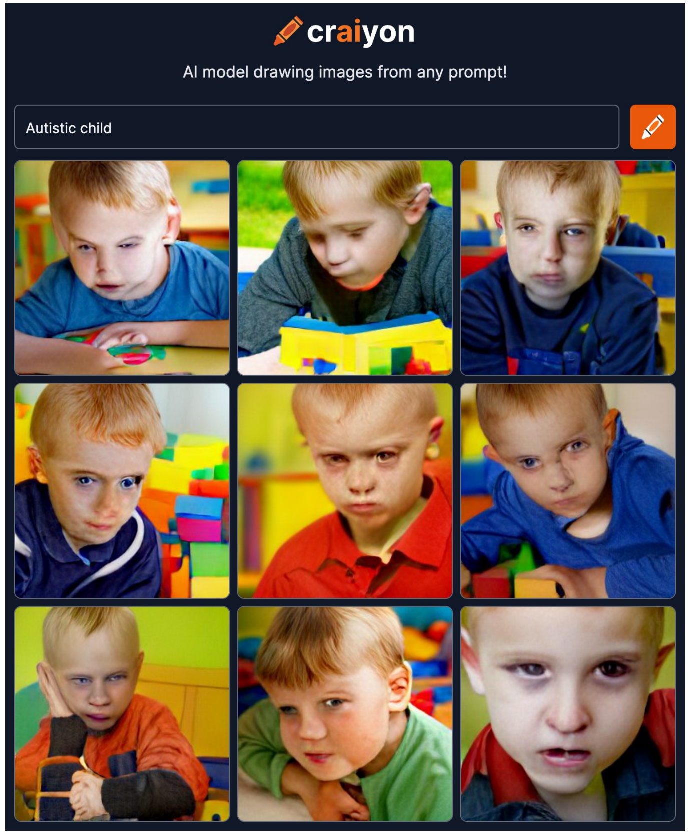


Autistic girl

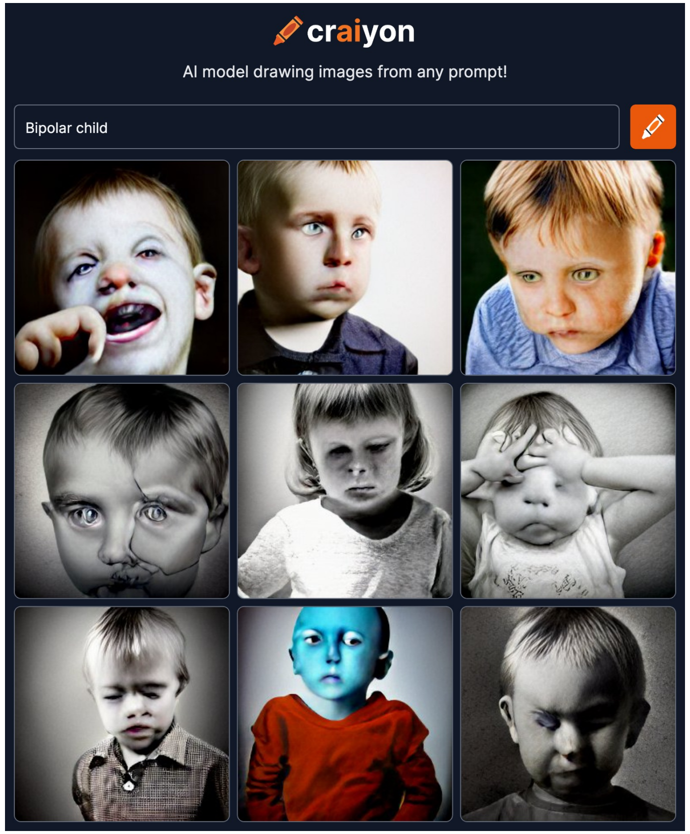
# Study case: neurodivergence



Neurodivergent child



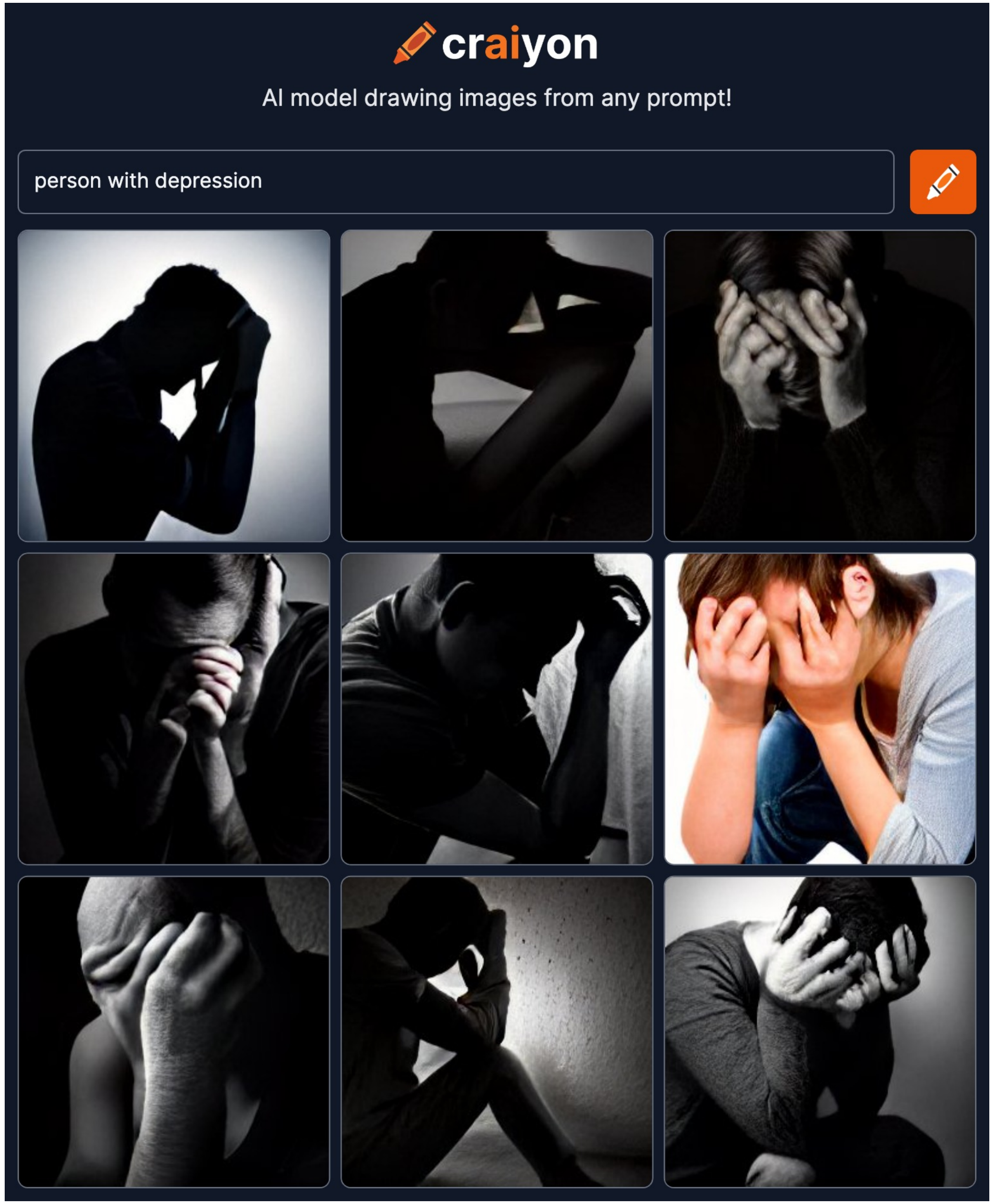
Autistic child



Bipolar child

# Study case: depression

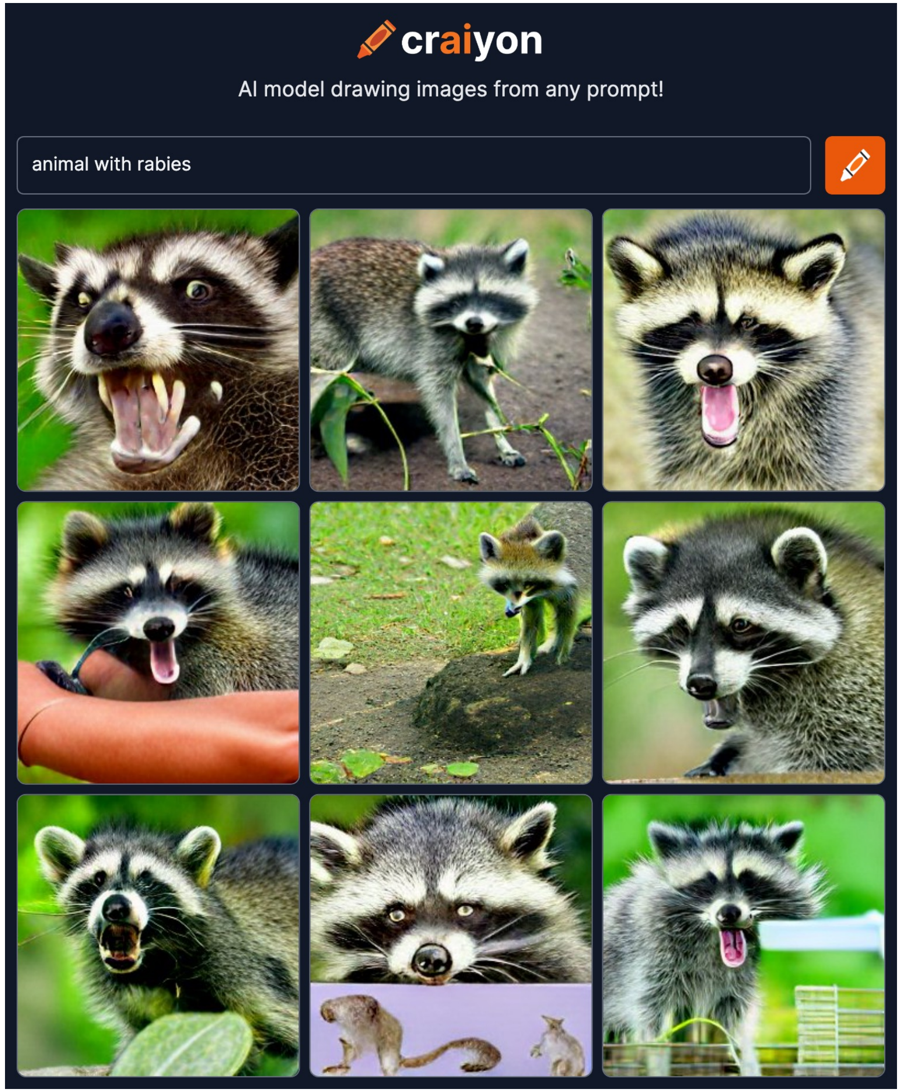
- All man
- All young
- Question: Does information, that not only youg adults have depression, is present in our collective imagination?
  - It is something totally different that we have article raising awarness and we kept some information in collective consciousness



person with depression

# Study case: rabies

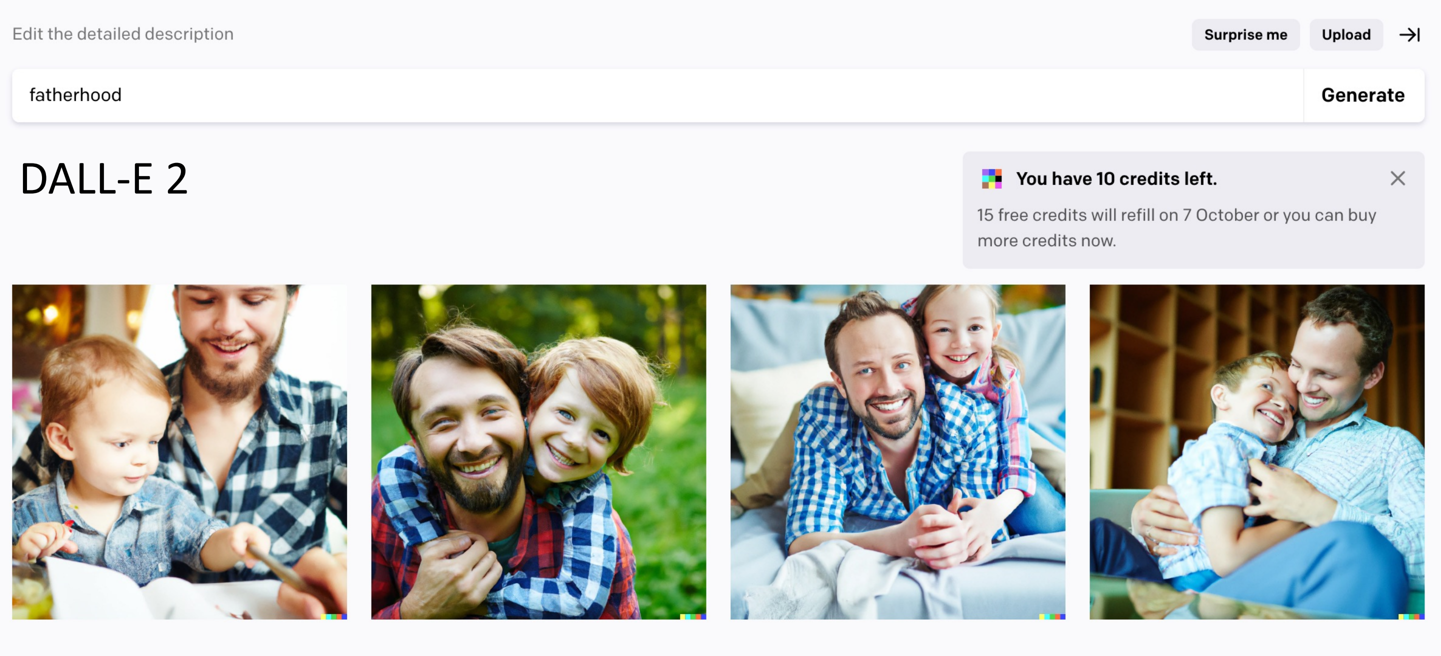
- We all know that dogs can transmit rabies. Americans know that raccoons can transmit rabies.
- All mamals can transmit rabies and lack of this knownledge can be deadly.



animal with rabies

# Study case: fatherhood

- Notice numer of children and their age.



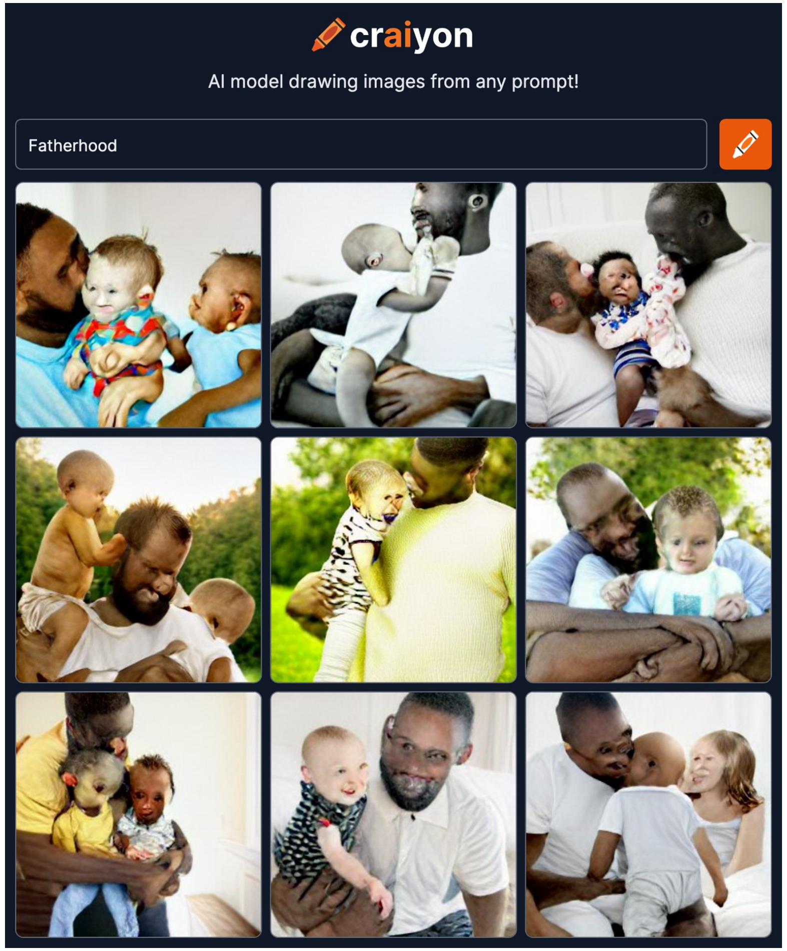
Edit the detailed description

fatherhood

DALL-E 2

You have 10 credits left.

15 free credits will refill on 7 October or you can buy more credits now.



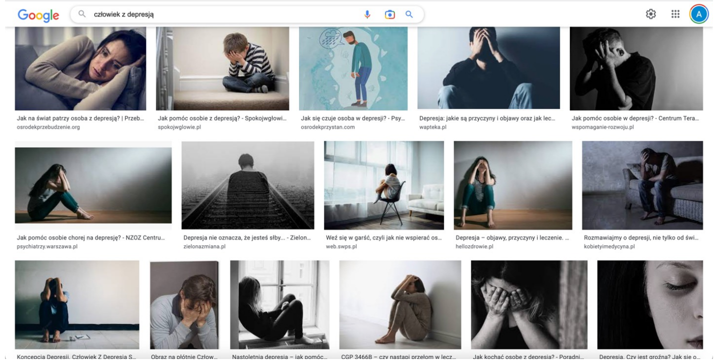
AI model drawing images from any prompt!

Fatherhood

# So maybe instead of AI we will use google images? Study case: depresion

- Google image in top 15 images (3 rows)
  - eng. no children (pl. 2 children)
  - Gender:
    - eng. 9 women (pl. 8 woman)
    - eng. 6 man  (pl. 5 man)
    - eng. 2 unable to indentify (pl. 0)
- Craiyon.com:
  - 8 man
  - 1 possibly women
- It seems that depression among children is not considered as a common issue in english internet, contrary to polish one

Question: Who is most likely to pay for the treatment?



człowiek z depresją

Jak na świat patrzy osoba z depresją? | Przeb...
osrodekprzebudzenie.org

Jak pomóc osobie z depresją? - Spokojwgłowi...
spokojwglowie.pl

Jak się czuje osoba w depresji? - Psy...
osrodekprzystan.com

Depresja: jakie są przyczyny i objawy oraz jak lec...
wapteka.pl

Jak pomóc osobie w depresji? - Centrum Tera...
wspomaganie-rozwoju.pl

Jak pomóc osobie chorej na depresję? - NZOZ Centru...
psychiatrzy.warszawa.pl

Depresja nie oznacza, że jesteś słby... - Zielon...
zielonazmiana.pl

Weź się w garść, czyli jak nie wspierać os...
web.swps.pl

Depresja – objawy, przyczyny i leczenie. ...
hellozdrowie.pl

Rozmawiajmy o depresji, nie tylko od świ...
kobietyimedycyna.pl

Koncepcja Depresji. Człowiek Z Depresja S...

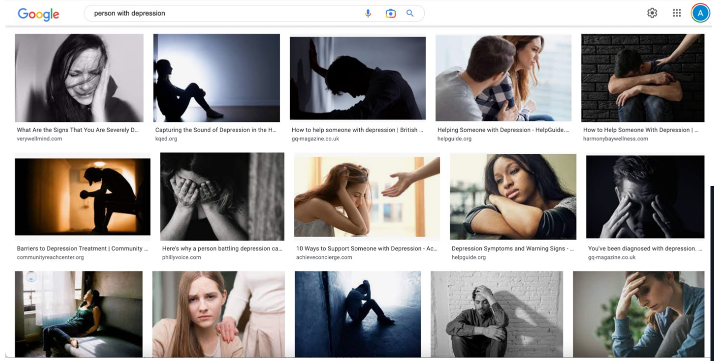
Obraz na płótnie Człow...

Nastoletnia depresja – jak pomóc...

CGP 3466B – czy nastąpi przełom w lecz...

Jak kochać osobę z depresją? - Poradni...

Depresja. Czy jest groźna? Jak się o...



person with depression

What Are the Signs That You Are Severely D...
verywellmind.com

Capturing the Sound of Depression in the H...
kqed.org

How to help someone with depression | British ...
gq-magazine.co.uk

Helping Someone with Depression - HelpGuide....
helpguide.org

How to Help Someone With Depression | ...
harmonybaywellness.com

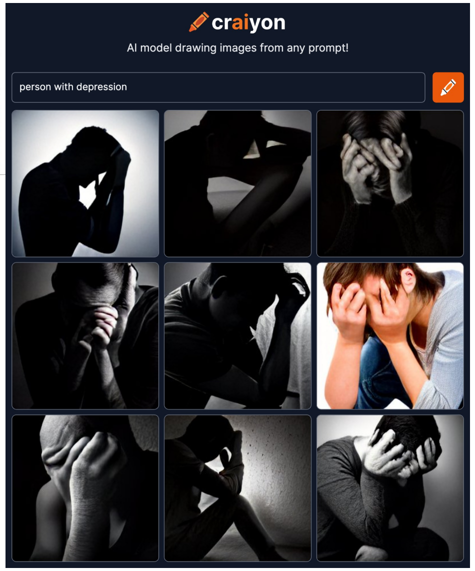
Barriers to Depression Treatment | Community ...
communityreachcenter.org

Here's why a person battling depression ca...
phillyvoice.com

10 Ways to Support Someone with Depression - Ac...
achieveconcierge.com

Depression Symptoms and Warning Signs - ...
helpguide.org

You've been diagnosed with depression. ...
gq-magazine.co.uk



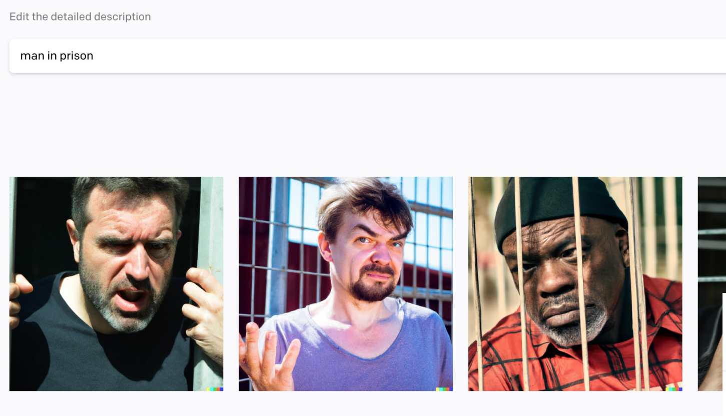
craiyon

AI model drawing images from any prompt!

person with depression

# Studing colective imagination in other domains

- We can also focus on media biases
- We are not limited to medicine



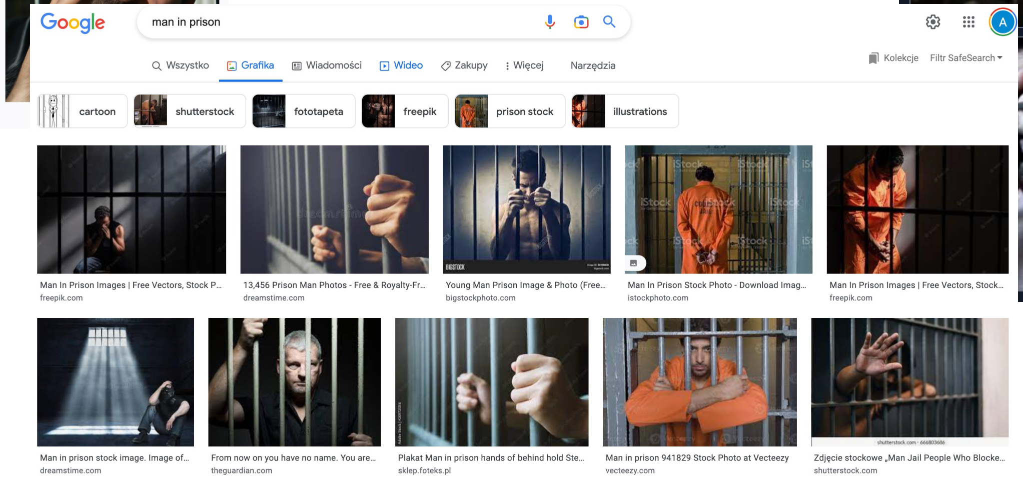
Edit the detailed description

Surprise me | Upload

man in prison

Generate

DALL-E 2



man in prison

Man In Prison Images | Free Vectors, Stock P...
freepik.com

13,456 Prison Man Photos - Free & Royalty-Fr...
dreamstime.com

Young Man Prison Image & Photo (Free...
bigstockphoto.com

Man In Prison Stock Photo - Download Imag...
istockphoto.com

Man In Prison Images | Free Vectors, Stock...
freepik.com

Man in prison stock image. Image of...
dreamstime.com

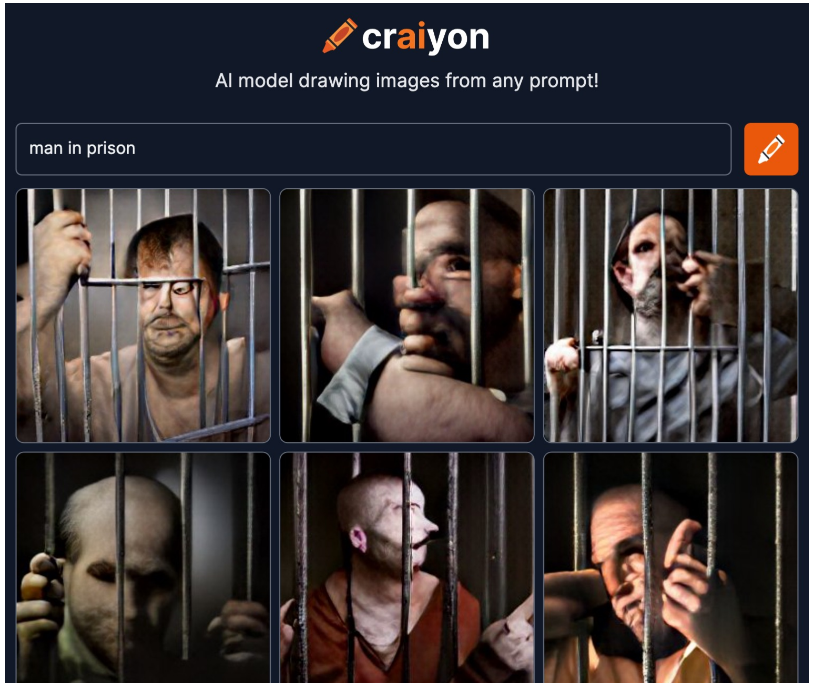
From now on you have no name. You are...
theguardian.com

Plakat Man in prison hands of behind hold Ste...
sklep.foteks.pl

Man in prison 941829 Stock Photo at Vecteezy
vecteezy.com

Zdjęcie stockowe „Man Jail People Who Blocke...
shutterstock.com

Google Images



AI model drawing images from any prompt!

man in prison

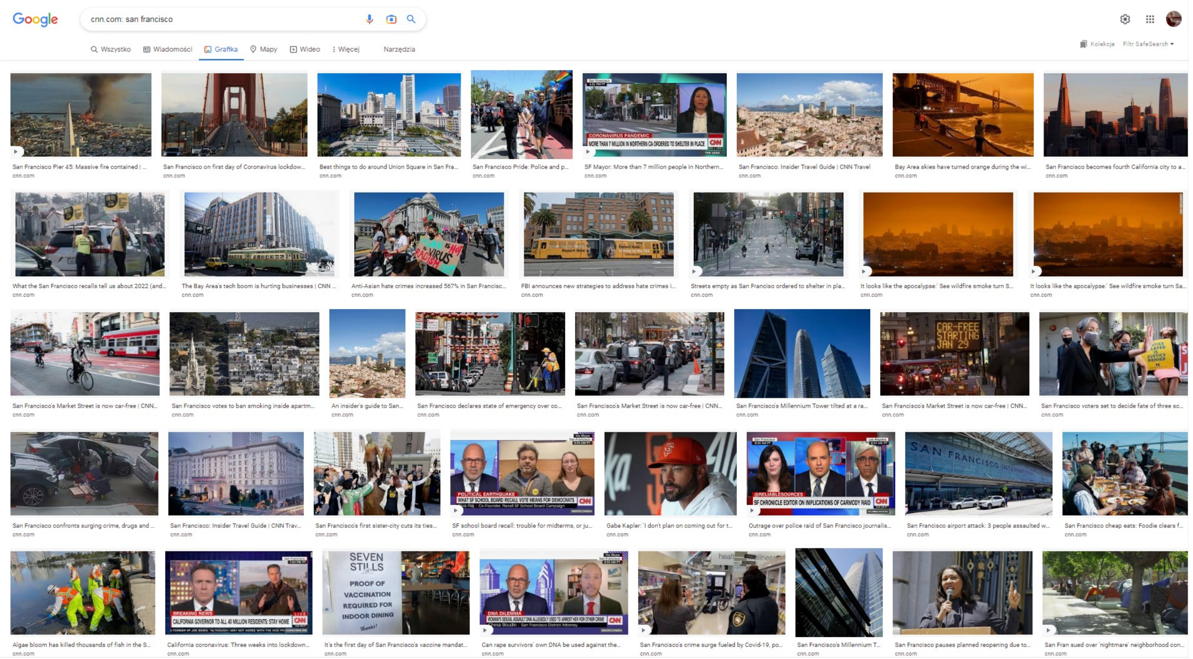
craiyon.com

# How to focus on certain segment of the society?

- If we fine-tune a model for certain media, webpages, resources that certain social group uses than we can possibly study

FOX
NEWS
channel

# How to focus on certain segment of the society?



cnn.com: san francisco

San Francisco Pier 45: Massive fire contained | ...
cnn.com

San Francisco on first day of Coronavirus lockdow...
cnn.com

Best things to do around Union Square in San Fra...
cnn.com

San Francisco Pride: Police and p...
cnn.com

SF Mayor: More than 7 million people in Northern...
cnn.com

San Francisco: Insider Travel Guide | CNN Travel
cnn.com

Bay Area skies have turned orange during the wi...
cnn.com

San Francisco becomes fourth California city to a...
cnn.com

What the San Francisco recalls tell us about 2022 (and...
cnn.com

The Bay Area's tech boom is hurting businesses | CNN ...
cnn.com

Anti-Asian hate crimes increased 567% in San Francisc...
cnn.com

FBI announces new strategies to address hate crimes i...
cnn.com

Streets empty as San Franciso ordered to shelter in pla...
cnn.com

It looks like the apocalypse:' See wildfire smoke turn S...
cnn.com

It looks like the apocalypse:' See wildfire smoke turn Sa...
cnn.com

San Francisco's Market Street is now car-free | CNN...
cnn.com

San Francisco votes to ban smoking inside apartm...
cnn.com

An insider's guide to San...
cnn.com

San Francisco declares state of emergency over co...
cnn.com

San Francisco's Market Street is now car-free | CNN...
cnn.com

San Francisco's Millennium Tower tilted at a ra...
cnn.com

San Francisco's Market Street is now car-free | CNN...
cnn.com

San Francisco voters set to decide fate of three sc...
cnn.com

San Francisco confronts surging crime, drugs and ...
cnn.com

San Francisco: Insider Travel Guide | CNN Trav...
cnn.com

San Francisco's first sister-city cuts its ties...
cnn.com

SF school board recall: trouble for midterms, or ju...
cnn.com

Gabe Kapler: 'I don't plan on coming out for t...
cnn.com

Outrage over police raid of San Francisco journalis...
cnn.com

San Francisco airport attack: 3 people assaulted w...
cnn.com

San Francisco cheap eats: Foodie clears f...
cnn.com

Algae bloom has killed thousands of fish in the S...
cnn.com

California coronavirus: Three weeks into lockdown...
cnn.com

It's the first day of San Francisco's vaccine mandat...
cnn.com

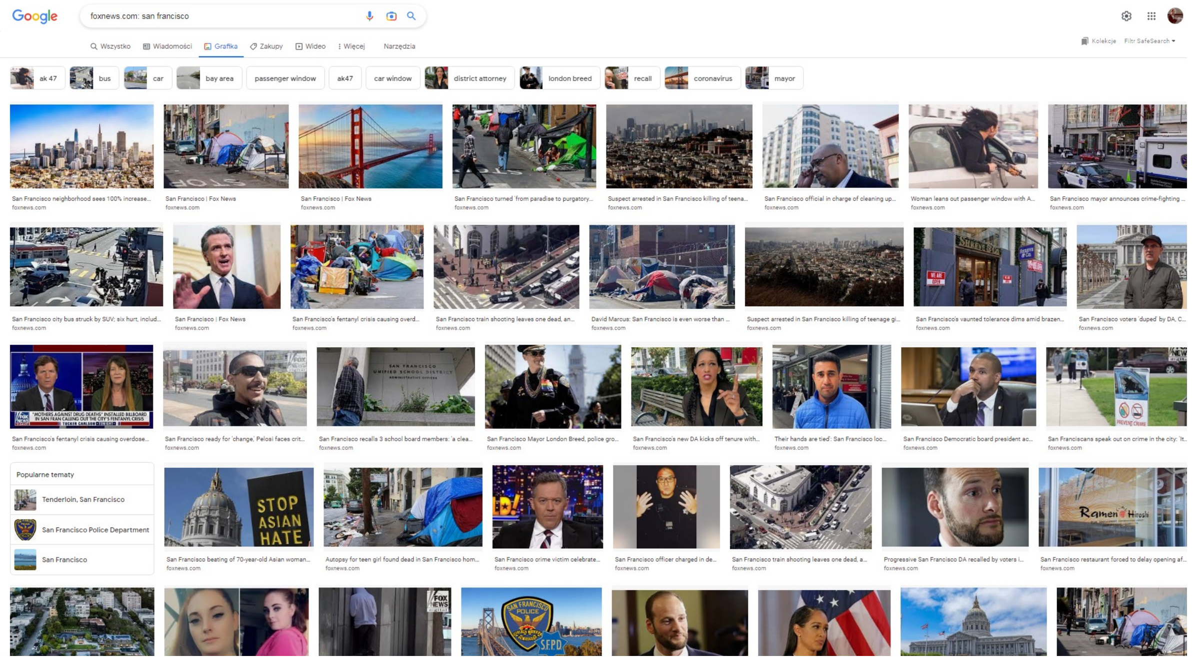
Can rape survivors' own DNA be used against the...
cnn.com

San Francisco's crime surge fueled by Covid-19, po...
cnn.com

San Francisco's Millennium T...
cnn.com

San Francisco pauses planned reopening due to...
cnn.com

San Fran sued over 'nightmare' neighborhood con...
cnn.com



foxnews.com: san francisco

San Francisco neighborhood sees 100% increase...
foxnews.com

San Francisco | Fox News
foxnews.com

San Francisco | Fox News
foxnews.com

San Francisco turned 'from paradise to purgatory...
foxnews.com

Suspect arrested in San Francisco killing of teena...
foxnews.com

San Francisco official in charge of cleaning up...
foxnews.com

Woman leans out passenger window with A...
foxnews.com

San Francisco mayor announces crime-fighting ...
foxnews.com

San Francisco city bus struck by SUV; six hurt, includ...
foxnews.com

San Francisco | Fox News
foxnews.com

San Francisco's fentanyl crisis causing overd...
foxnews.com

San Francisco train shooting leaves one dead, an...
foxnews.com

David Marcus: San Francisco is even worse than ...
foxnews.com

Suspect arrested in San Francisco killing of teenage gi...
foxnews.com

San Francisco's vaunted tolerance dims amid brazen...
foxnews.com

San Francisco voters 'duped' by DA, C...
foxnews.com

San Francisco's fentanyl crisis causing overdose...
foxnews.com

San Francisco ready for 'change,' Pelosi faces crit...
foxnews.com

San Francisco recalls 3 school board members: 'a clea...
foxnews.com

San Francisco Mayor London Breed, police gro...
foxnews.com

San Francisco's new DA kicks off tenure with...
foxnews.com

Their hands are tied': San Francisco loc...
foxnews.com

San Francisco Democratic board president ac...
foxnews.com

San Franciscans speak out on crime in the city: 'It...
foxnews.com

Popularne tematy
- Tenderloin, San Francisco
- San Francisco Police Department
- San Francisco

San Francisco beating of 70-year-old Asian woman...
foxnews.com

Autopsy for teen girl found dead in San Francisco hom...
foxnews.com

San Francisco crime victim celebrate...
foxnews.com

San Francisco officer charged in de...
foxnews.com

San Francisco train shooting leaves one dead, a...
foxnews.com

Progressive San Francisco DA recalled by voters i...
foxnews.com

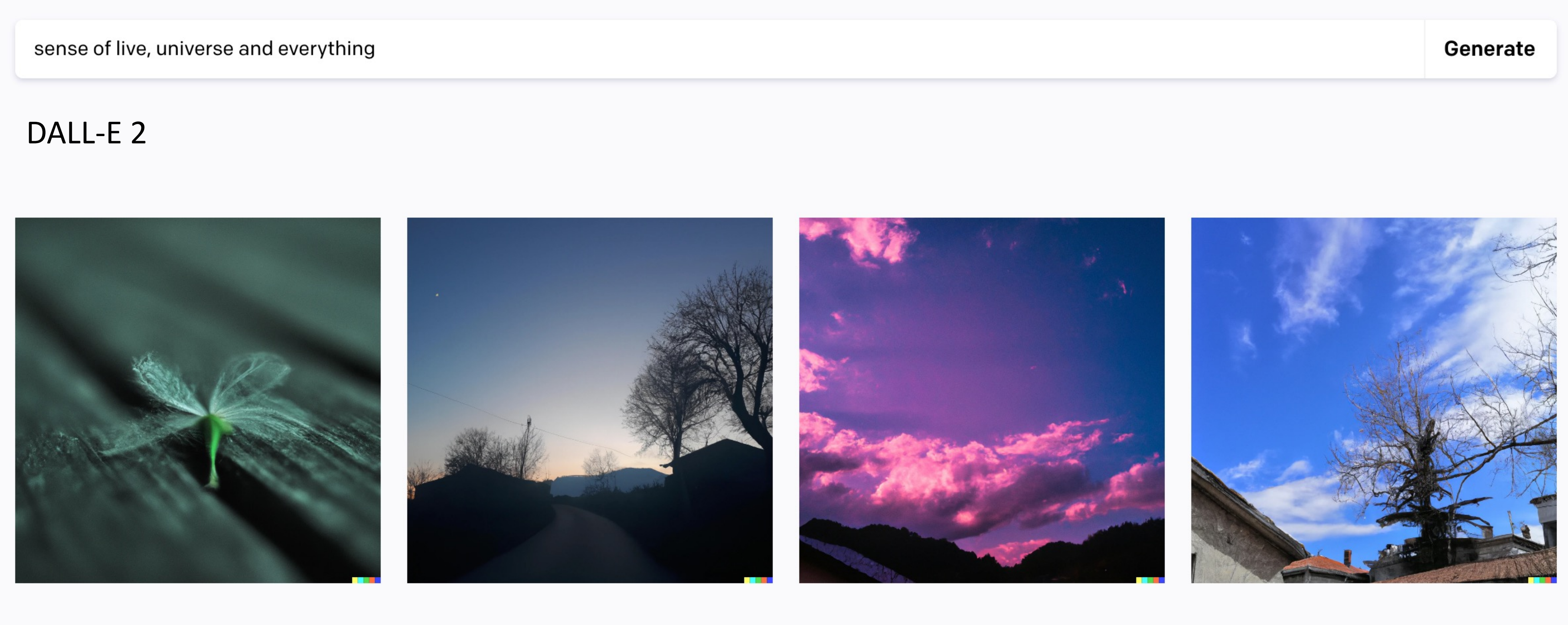
San Francisco restaurant forced to delay opening af...
foxnews.com

CNN

FOX NEWS channel

# Ultimate question of live, universe and everything

# Ultimate question of live, universe and everything

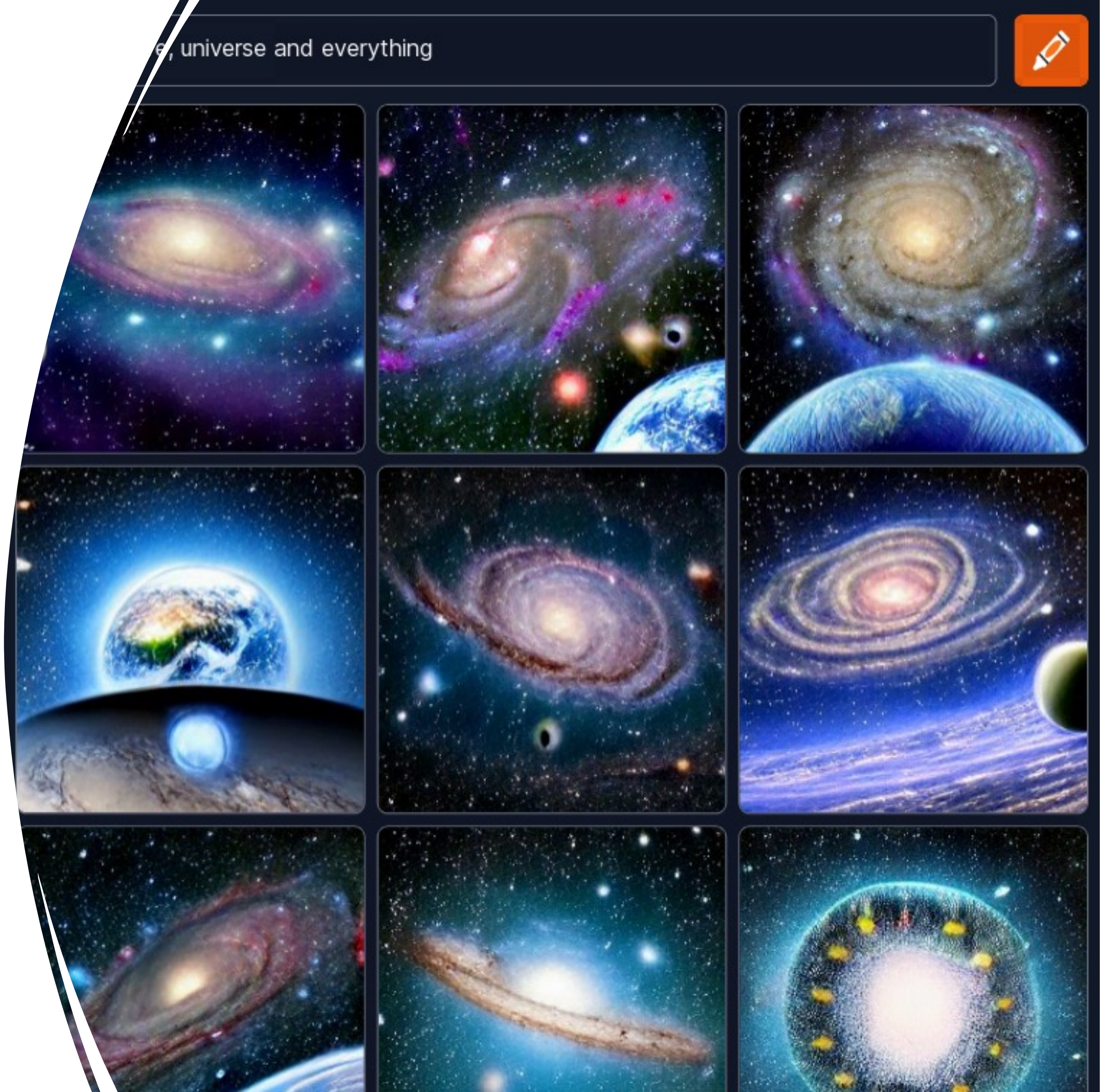


sense of live, universe and everything

DALL-E 2

# Summary

- AI encodes some biases that are present in training data (sourced for media)
- And media are source of biases to people
- Using text-to-image models we can probe some of the biases connected to medical condisions
- But we are not limited to them
- It is possible to target also certain media segments that are connected with studied population



e, universe and everything