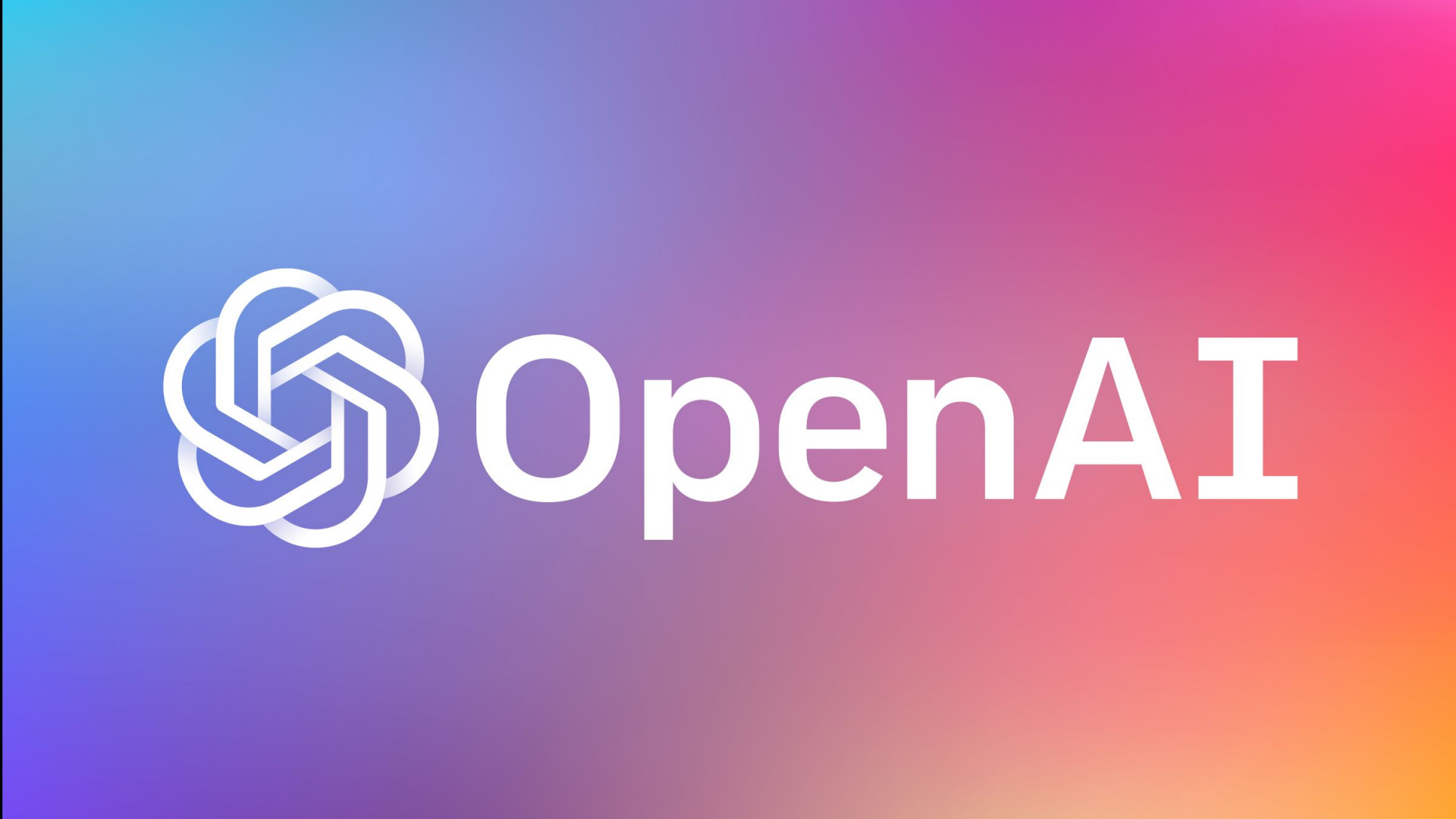
DALL·E

DALL·E: Creating Images from Text

We've trained a neural network called DALL·E that creates images from text captions for a wide range of concepts expressible in natural language.

January 5, 2021
27 minute read


OpenAI

TEXT PROMPT

an illustration of a baby daikon radish in a tutu walking a dog

AI-GENERATED IMAGES

Edit prompt or view more images↓

TEXT PROMPT

an armchair in the shape of an avocado [...]

AI-GENERATED IMAGES

Edit prompt or view more images↓

Table of contents

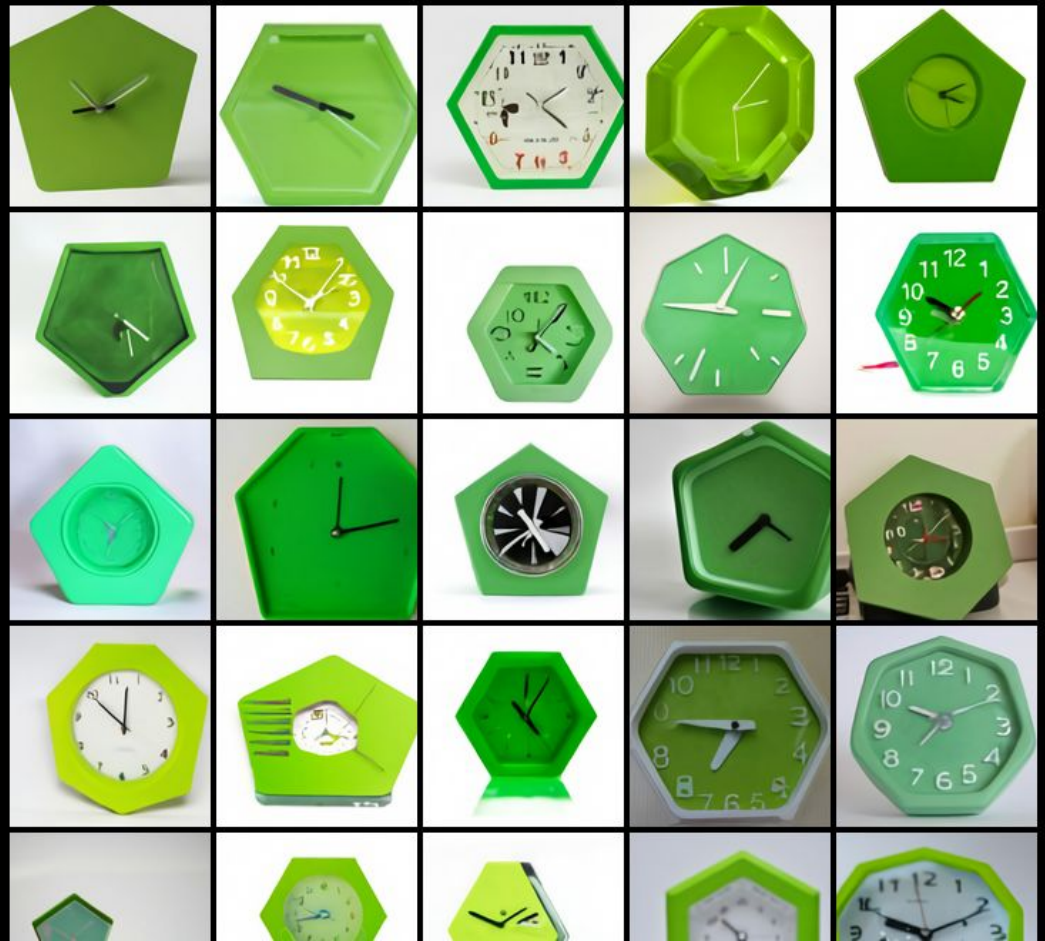
Capabilities

Controlling attributes

TEXT PROMPT

a pentagonal green clock. a green clock in the shape of a pentagon.

AI-GENERATED IMAGES

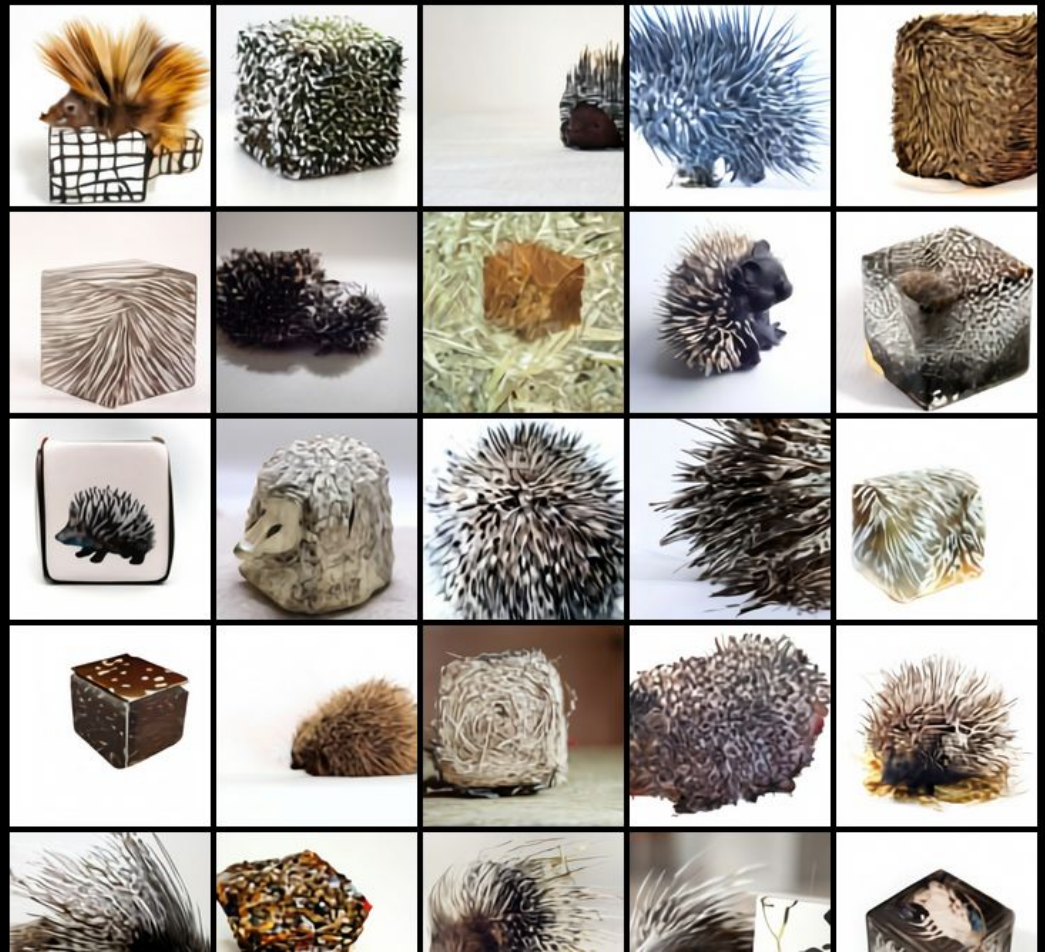


We find that DALL-E can
in polygonal shapes tha
to occur in the real worl
as "picture frame" and "
reliably draw the object
shapes except heptagon
as "manhole cover" and
success rate for more u
"pentagon," is considera

For several of the visual
that repeating the capti
alternative phrasings, in
of the results.

TEXT PROMPT a cube made of porcupine. a cube with the texture of a porcupine.

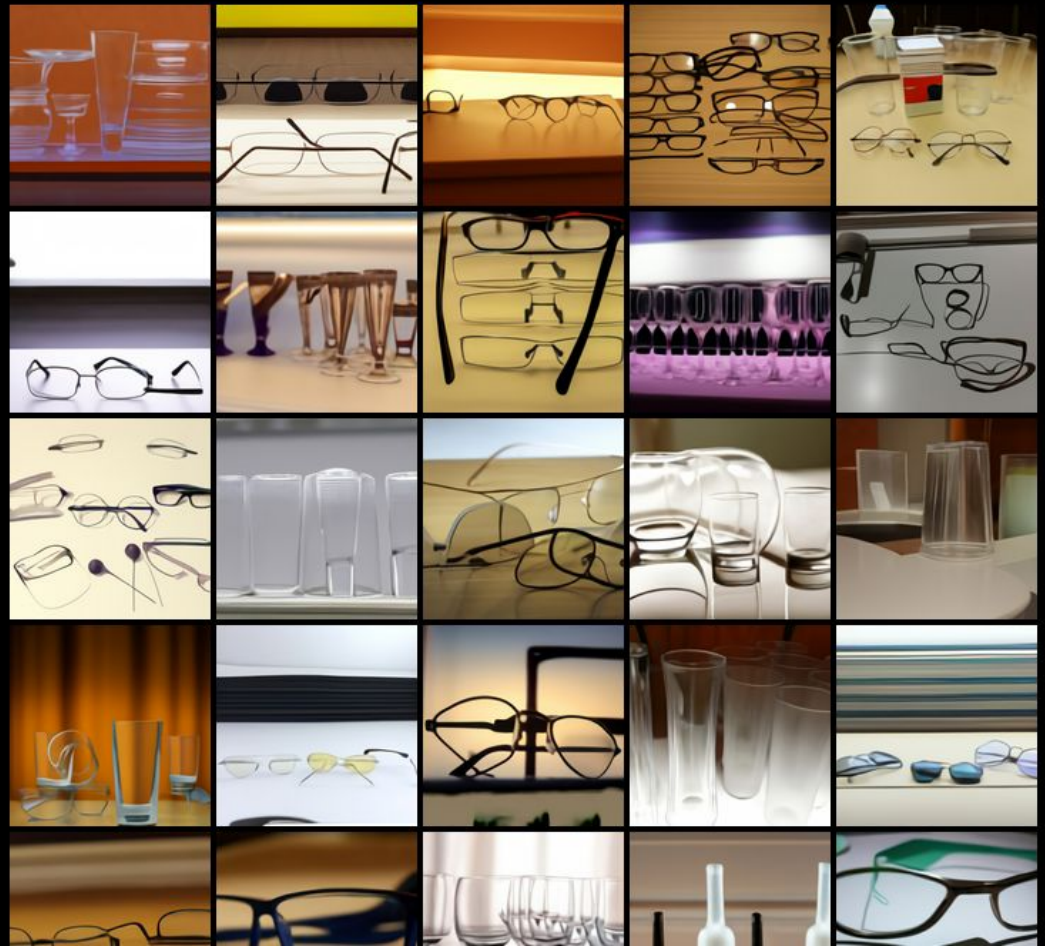
AI-GENERATED IMAGES



We find that DALL-E can
various plants, animals,
three dimensional solid
visual, we find that repe
alternative phrasing imp
the results.

TEXT PROMPT a collection of glasses is sitting on a table

AI-GENERATED IMAGES



We find that DALL·E is a
copies of an object whe
is unable to reliably cou
prompted to draw noun
multiple meanings, such
and "cups" it sometime
interpretations, depend
that is used.

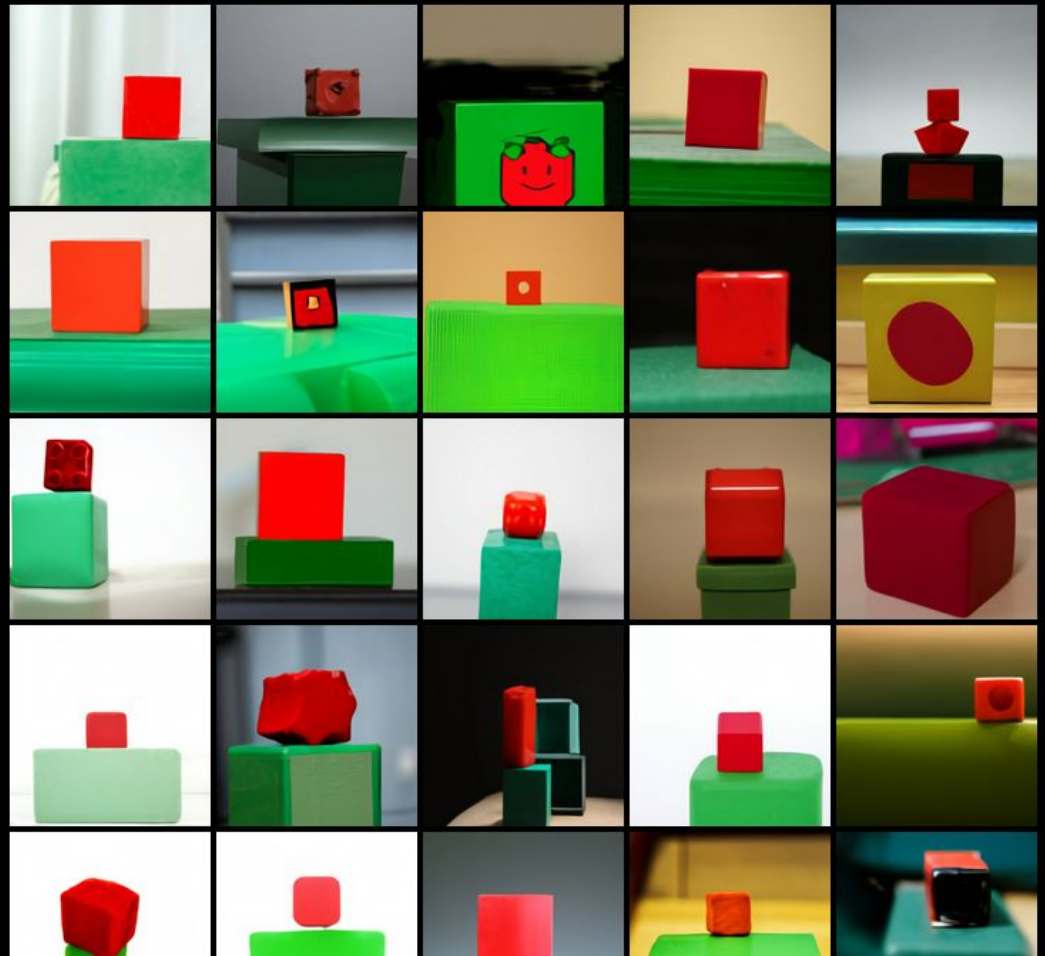
Capabilities

Controlling attributes

Drawing multiple objects

TEXT PROMPT a small red block sitting on a large green block

AI-GENERATED IMAGES

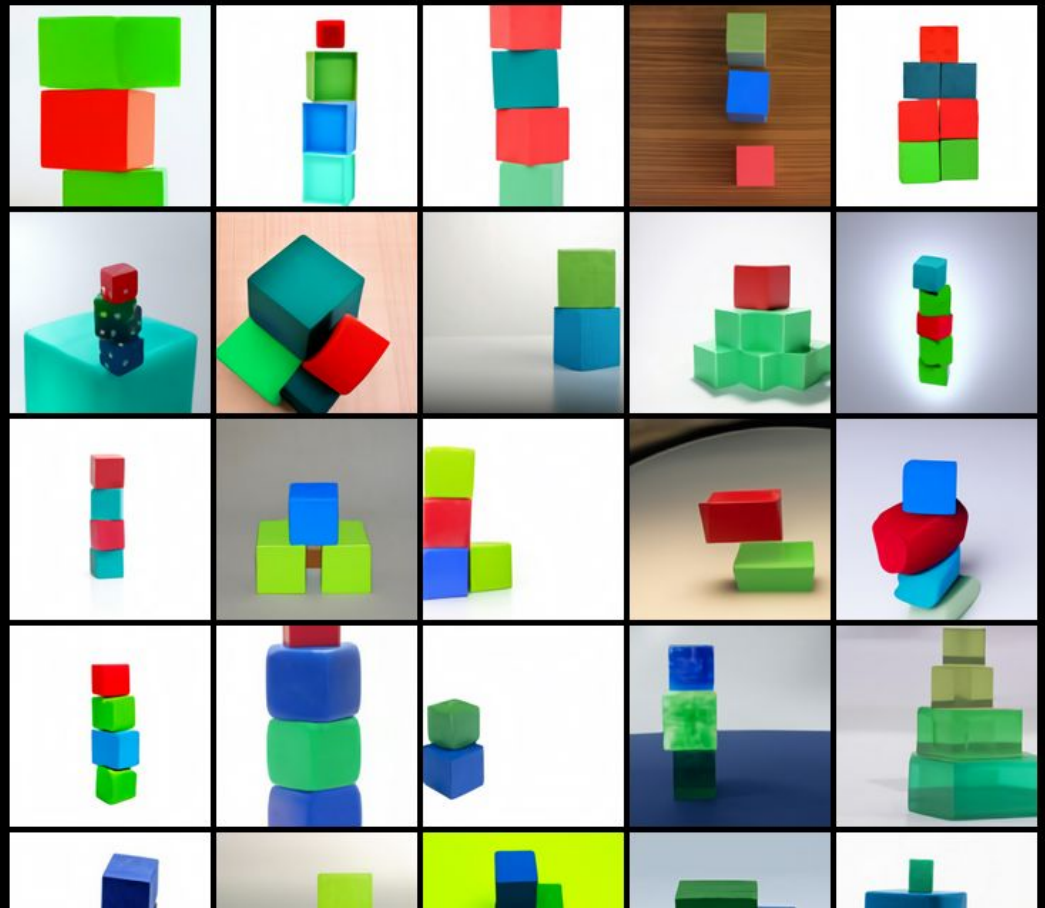


We find that DALL·E cor
types of relative positio
choices "sitting on" and
sometimes appear to w
"standing behind," "star
"standing right of" do n
lower success rate whe
object sitting on top of a
compared to the other v

TEXT PROMPT

a stack of 3 cubes. a red cube is on the top, sitting on a green cube. the green cube is in the middle, sitting on a blue cube. the blue cube is on the bottom.

AI-GENERATED IMAGES



We find that DALL-E typ
image with one or two o
correct colors. However,
each setting tend to hav
colored precisely as spe

TEXT PROMPT

an emoji of a baby penguin wearing a blue hat, red gloves, green shirt, and yellow pants

AI-GENERATED IMAGES

We find that DALL-E typ
image with two or three
having the correct color
the samples for each se
articles of clothing with

Capabilities

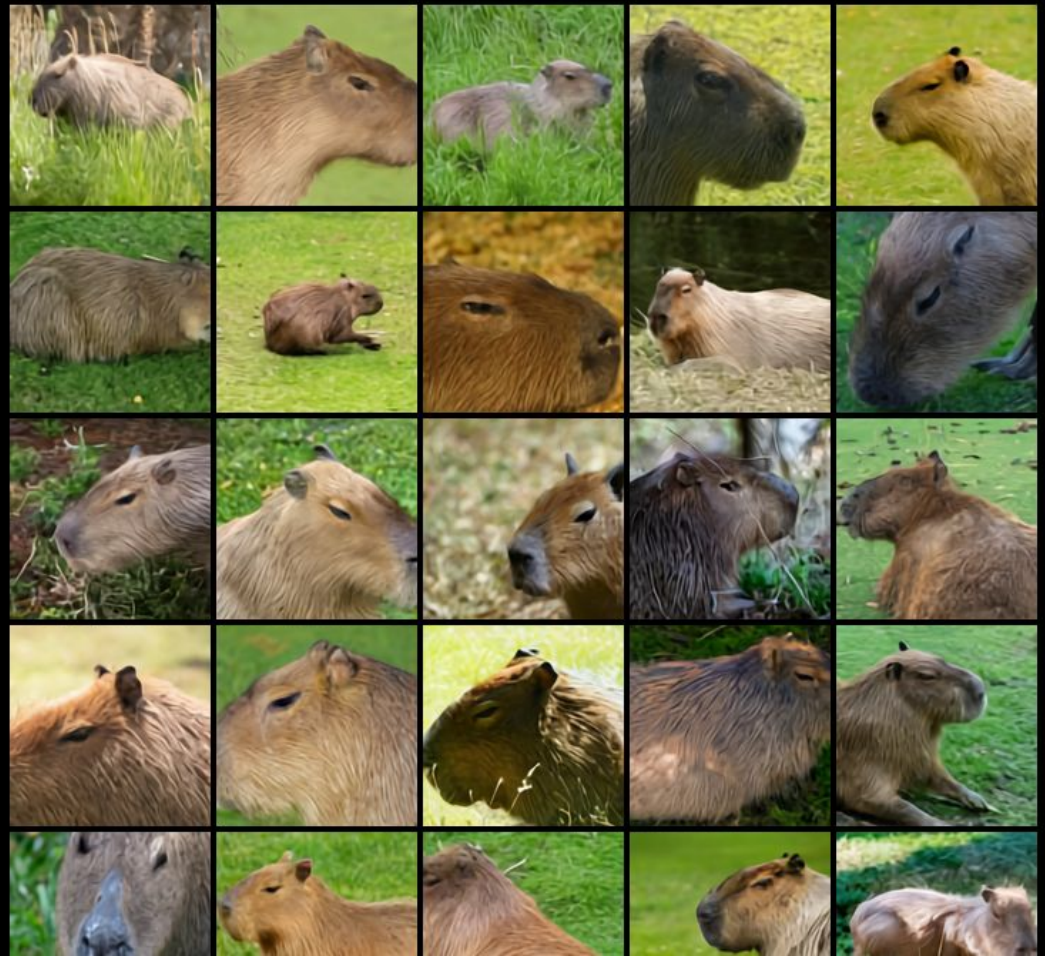
Controlling attributes

Drawing multiple objects

Perspective and 3 - dimensionality

TEXT PROMPT an extreme close-up view of a capybara sitting in a field

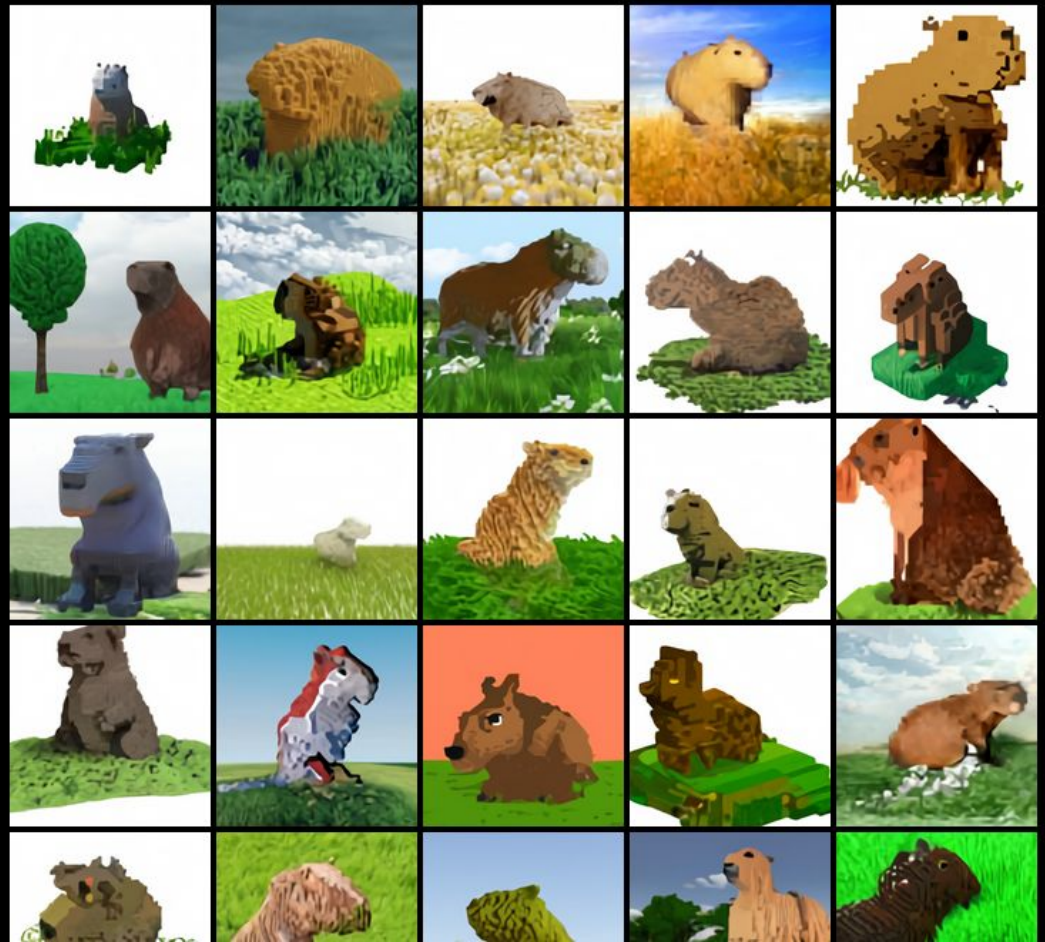
AI-GENERATED IMAGES



We find that DALL·E can
animals in a variety of d
these views, such as "ae
view," require knowledg
appearance from unusu
as "extreme close-up vi
the fine-grained details
fur.

TEXT PROMPT a capybara made of voxels sitting in a field

AI-GENERATED IMAGES

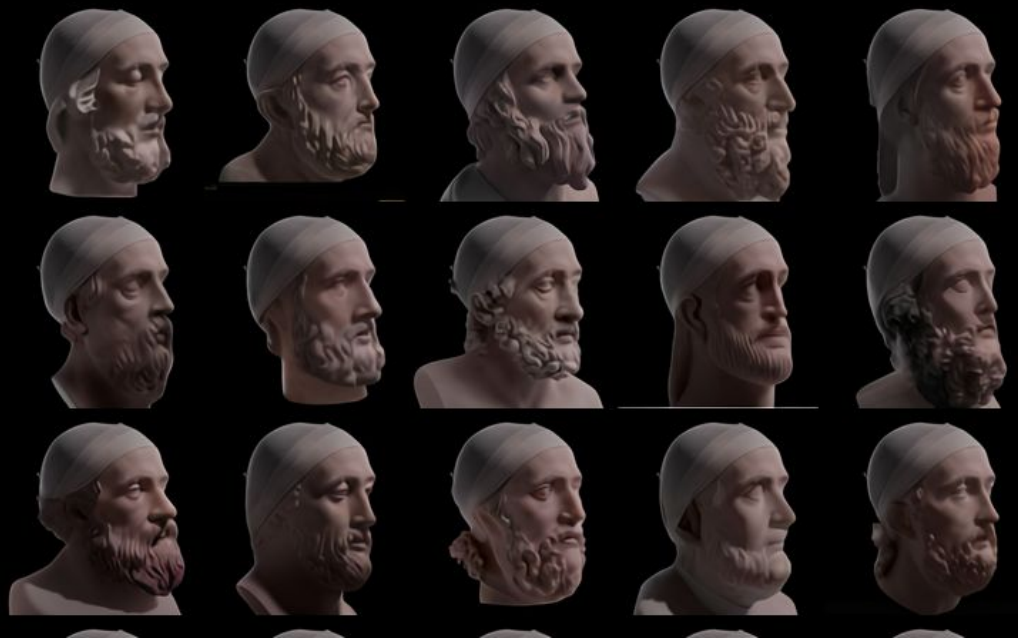


We find that DALL·E is o
surface of each of the a
chosen 3D style, such a
"made of voxels," and re
plausible shading depe
the sun. The "x-ray" styl
reliably, but it shows tha
orient the bones within
(though not anatomical
configurations.

TEXT PROMPT a photograph of a bust of homer

IMAGE PROMPT

AI-GENERATED IMAGES



We prompt DALL-E with
describing a well-known
region of an image show
particular angle. Then, w
complete the remaining
this contextual informat
repeatedly, each time ro
more degrees, and find
recover smooth animati
known figures, with eac
precise specification of
lighting.

TEXT PROMPT a photograph of a bust of homer

IMAGE PROMPT

AI-GENERATED IMAGES

We prompt DALL-E with
describing a well-known
region of an image show
particular angle. Then, w
complete the remaining
this contextual informat
repeatedly, each time ro
more degrees, and find
recover smooth animati
known figures, with eac
precise specification of
lighting.

TEXT PROMPT

a plain white cube looking at its own reflection in a mirror. a plain white cube gazing at itself in a mirror.

IMAGE PROMPT

AI-GENERATED IMAGES

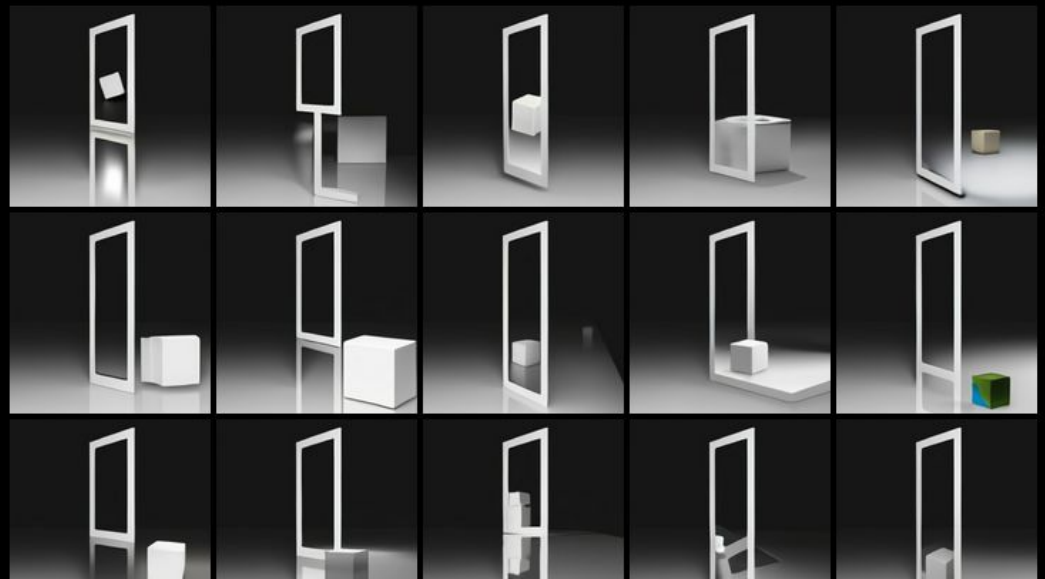
Similar to what was don
DALL-E to complete the
a sequence of frames, e
mirror and reflective flo
the mirror usually resem
of it, it often does not re
physically correct way. E
of an object drawn on a
typically more plausible

TEXT PROMPT

a plain white cube looking at its own reflection in a mirror. a plain white cube gazing at itself in a mirror.

IMAGE PROMPT

AI-GENERATED IMAGES



Similar to what was don
DALL·E to complete the
a sequence of frames, e
mirror and reflective flo
the mirror usually resem
of it, it often does not re
physically correct way. E
of an object drawn on a
typically more plausible

Capabilities

Controlling attributes

Drawing multiple objects

Perspective and 3 - dimensionality

Internal and external structures

TEXT PROMPT a cross-section view of a walnut

AI-GENERATED IMAGES

We find that DALL·E is a
of several different kind

TEXT PROMPT a macro photograph of brain coral

AI-GENERATED IMAGES

We find that DALL-E is a
grained external details
of objects. These details
the object is viewed up

Capabilities

Controlling attributes

Drawing multiple objects

Perspective and 3 - dimensionality

Internal and external structures

Inferring contextual details

TEXT PROMPT a ... of a capybara sitting in a field at sunrise

AI-GENERATED IMAGES

painting

painting in pop art style

painting in cubist style

painting in surrealist style

painting in the style of van gogh

painting in the style of claude monet

drawing

charcoal drawing

crayon drawing

chalk drawing

We find that DALL·E is a
scene in a variety of diff
adapt the lighting, shad
based on the time of da

TEXT PROMPT a stained glass window with an image of a blue strawberry

AI-GENERATED IMAGES

We find that DALL·E is a
representation of the ob
medium on which it is b
mural," "a soda can," an
must change how it dra
the angle and curvature
For "a stained glass win
it must alter the appear
how it usually appears.

TEXT PROMPT

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

AI-GENERATED IMAGES

We find that DALL·E is s
text and adapt the writi
which it appears. For ex
and "a license plate" ea
types of fonts, and "a ne
the sky" require the app
be changed.

Generally, the longer the
prompted to write, the l
We find that the succes
parts of the caption are
the success rate somet
sampling temperature f
decreased, although the
simpler and less realisti

Capabilities

Controlling attributes

Drawing multiple objects

Perspective and 3 - dimensionality

Internal and external structures

Inferring contextual details

Applications of preceding capabilities

TEXT PROMPT

a male mannequin dressed in an orange and black flannel shirt and black jeans

IMAGE PROMPT

AI-GENERATED IMAGES

We explore DALL·E's abi
mannequins in a variety
When prompted with tw
and white bomber jacke
black turtleneck sweate
a range of possibilities f
be used for the same ar

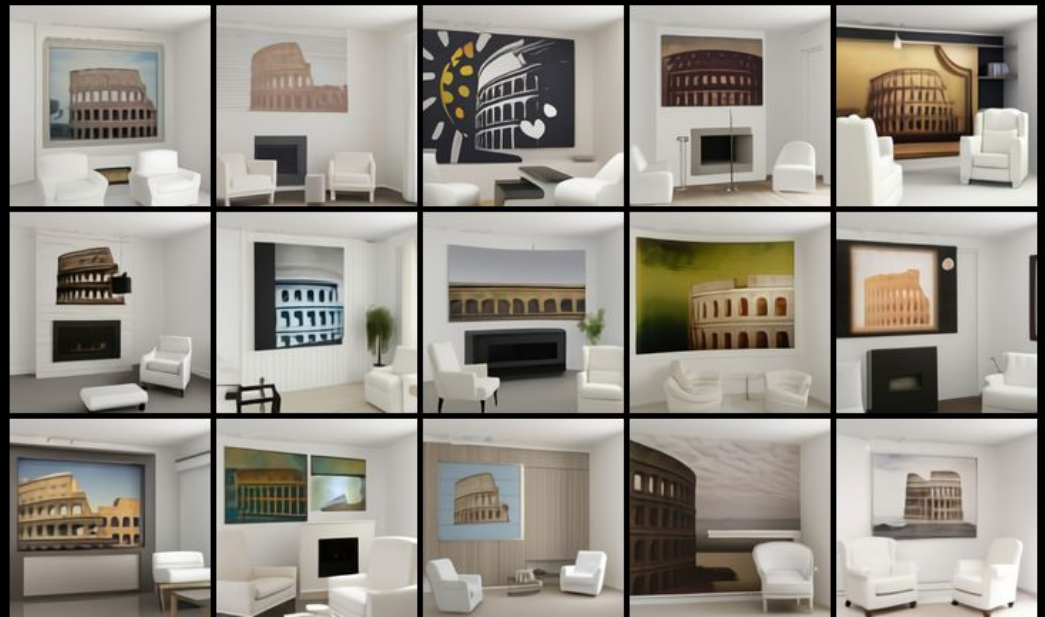
DALL·E also seems to o
common colors with oth
For example, when pron
"navy," DALL·E sometim
blue, or shades very clo
DALL·E sometimes con
of brown or brighter sha

TEXT PROMPT

a living room with two white armchairs and a painting of the collosseum. the painting is mounted above a modern fireplace.

IMAGE PROMPT

AI-GENERATED IMAGES



We explore DALL·E's abi
of rooms with several de
that it can generate pai
different subjects, inclu
such as "the colosseum
characters like "yoda." F
exhibits a variety of inte
painting is almost alway
DALL·E sometimes fails
the correct number of a

Capabilities

Controlling attributes

Drawing multiple objects

Perspective and 3 - dimensionality

Internal and external structures

Inferring contextual details

Applications of preceding capabilities

Combining unrelated concepts

TEXT PROMPT a snail made of harp. a snail with the texture of a harp.

AI-GENERATED IMAGES

We find that DALL·E can
synthesized from a varie
musical instruments, fo
items. While not always
DALL·E sometimes take
objects into considerati
how to combine them. F
prompted to draw "a sna
sometimes relates the p
spiral of the snail's shell

In a previous section, w
objects are introduced i
liable to confuse the ass
objects and their specif
see a different sort of fa
rather than binding som
specified concept (say,
(say, "a snail"), DALL·E j
separate items.

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

AI-GENERATED IMAGES

In the preceding visual,
ability to generate fanta
combining two unrelate
its ability to take inspira
idea while respecting th
designed, ideally produc
appears to be practicall
that prompting DALL·E
shape of," "in the form o
gives it the ability to do

When generating some
"an armchair in the shap
DALL·E appears to relat
avocado to the back of
the avocado to the cush
is susceptible to the sa
mentioned in the previo

Capabilities

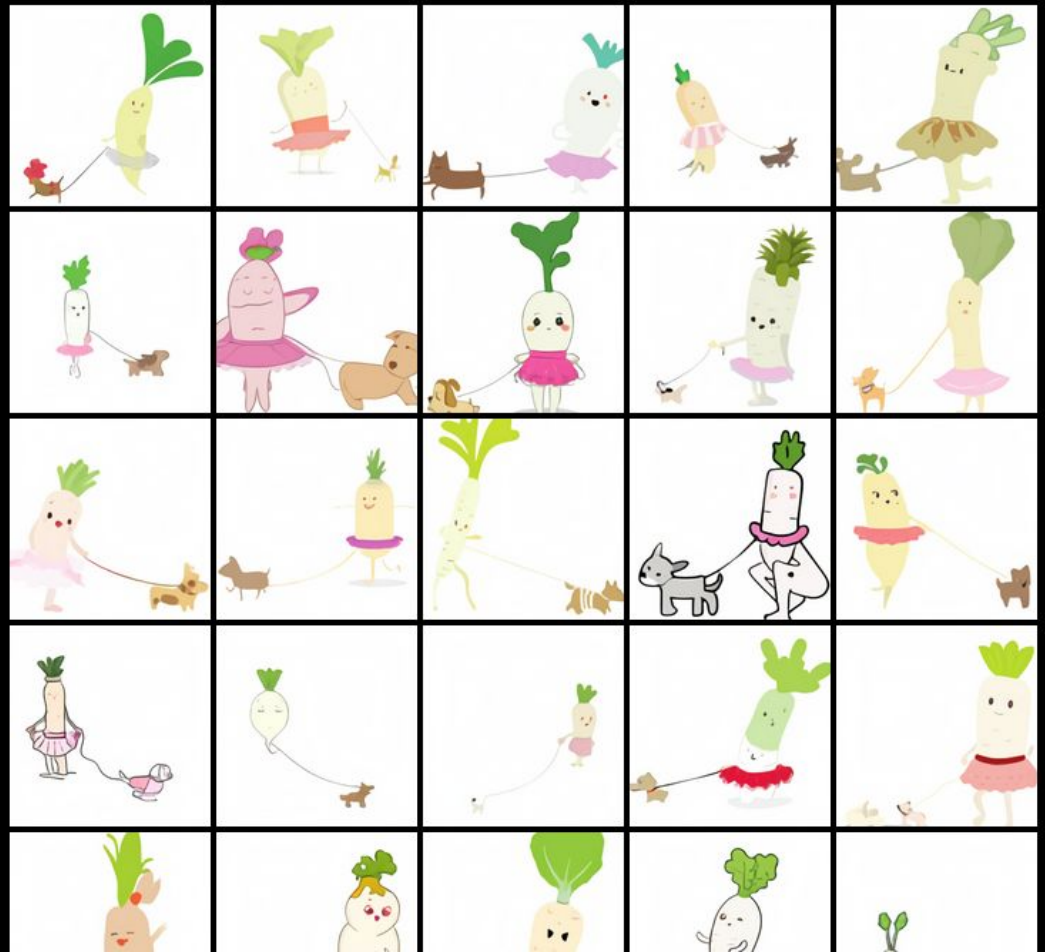
...

Combining unrelated concepts

Animal illustrations

TEXT PROMPT an illustration of a baby daikon radish in a tutu walking a dog

AI-GENERATED IMAGES



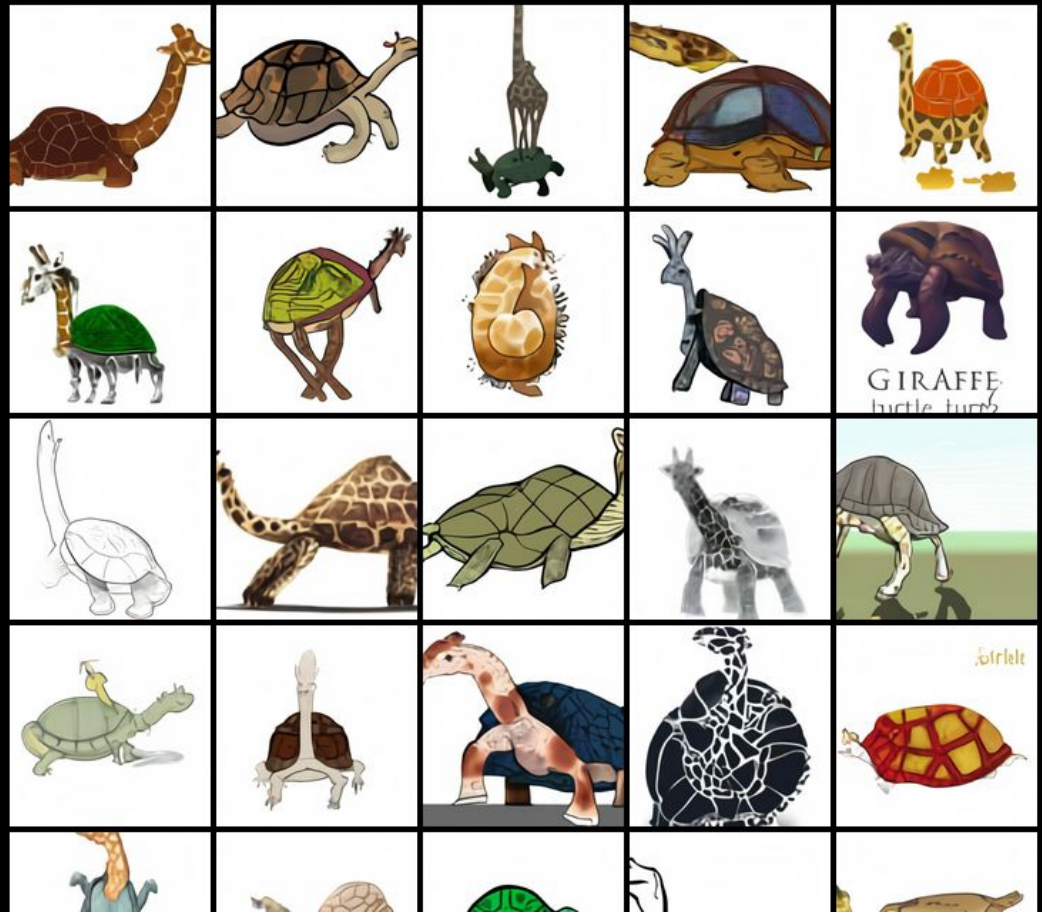
We find that DALL·E is s
transfer some human a
clothing to animals and
as food items. We inclu
"wielding a blue lightsa
ability to incorporate po

We find it interesting ho
body parts onto animals
asked to draw a daikon
sipping a latte, or riding
draws the kerchief, hand
locations.

TEXT PROMPT

a professional high quality illustration of a giraffe turtle chimera. a giraffe imitating a turtle. a giraffe made of turtle.

AI-GENERATED IMAGES



We find that DALL·E is s
combine distinct anima
include “pikachu” to exp
incorporate knowledge
“robot” to explore its ab
cyborgs. Generally, the
animal mentioned in the
dominant.

We also find that inserti
“professional high quali
and “emoji” sometimes
consistency of the resul

TEXT PROMPT a professional high quality emoji of a lovestruck cup of boba

AI-GENERATED IMAGES

We find that DALL·E is s
transfer some emojis to
objects, such as food ite
visual, we find that inse
"professional high quali
sometimes improves th
consistency of the resul

Capabilities

...

Combining unrelated concepts

Animal illustrations

zero shot visual reasoning

TEXT PROMPT

the exact same cat on the top as a sketch on the bottom

IMAGE PROMPT

AI-GENERATED IMAGES

We find that DALL·E is a
of image transformatio
with varying degrees of
straightforward ones, su
pink" and "photo reflect
tend to be the most relia
is often not copied or re
transformation "animal
view" requires DALL·E t
the animal in the photo,
with the appropriate de
reliably, and for several
only generates plausibl

TEXT PROMPT

the exact same teapot on the top with 'gpt' written on it on the bottom

IMAGE PROMPT

AI-GENERATED IMAGES

We find that DALL·E
different kinds of im
photos of teapots, w
reliability. Aside fron
color of the teapot (e
pattern (e.g., "with s
render text (e.g., "wi
map the letters onto
teapot in a plausible
reliability, it can also
size (for the "tiny" op
(for the "broken" opt

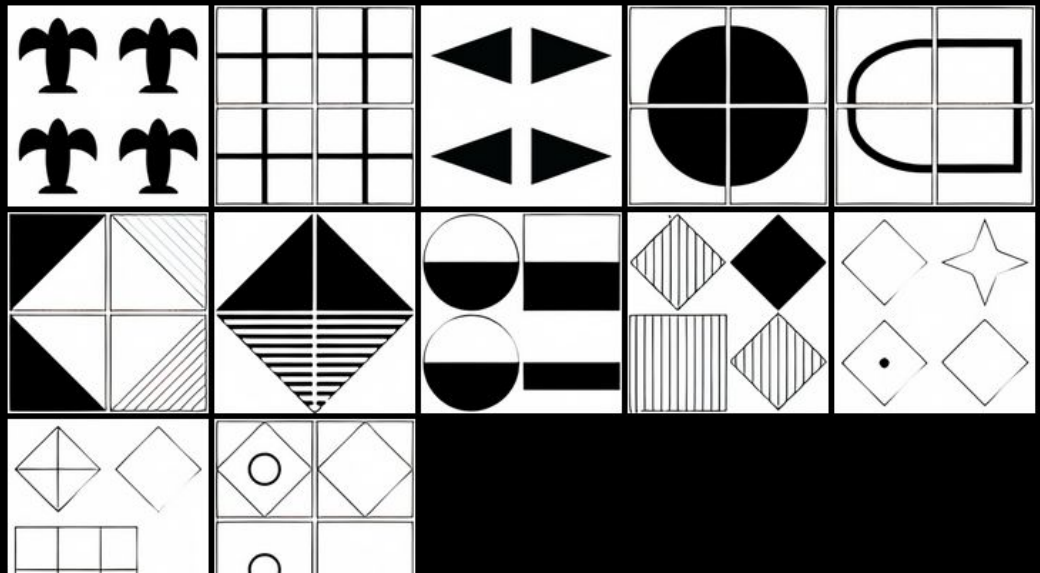
TEXT PROMPT

a sequence of geometric shapes. [Set D]

EXAMPLE IMAGE PROMPT

Set B ☐ Inverted

AI-GENERATED IMAGES



Rather than treating the
choice problem as origi
DALL·E to complete the
each image using argm
consider its completion
close visual match to th

DALL·E is often able to s
involve continuing simp
geometric reasoning, su
C. It is sometimes able
involve recognizing perm
boolean operations, suc
instances in set E tend
and DALL·E gets almost

For each of the sets, we
performance on both th
the images with the col

Capabilities

...

Combining unrelated concepts

Animal illustrations

zero shot visual reasoning

Geographic and temporal knowledge

TEXT PROMPT a photo of the food of china

AI-GENERATED IMAGES

We test DALL·E's unders
geographical facts, such
cuisines, and local wildl
successfully answers m
such as those involving
reflects superficial stere
"food" and "wildlife," as
the full diversity encoun

TEXT PROMPT a photo of alamo square, san francisco, from a street at night

AI-GENERATED IMAGES

We find that DALL·E is s
rendering semblances o
Francisco. For locations
such as San Francisco,
déjà vu—eerie simulacr
and cafes that remind u
locations that do not ex

TEXT PROMPT a photo of san francisco's golden gate bridge

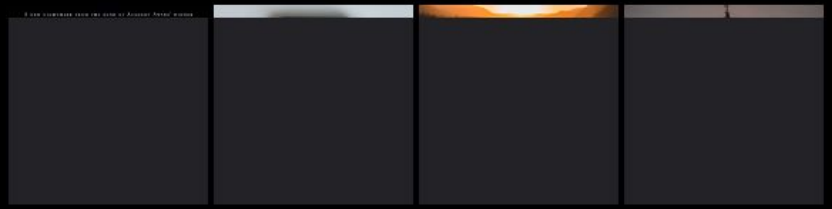
IMAGE PROMPTS

AI-GENERATED IMAGES

We can also prompt DA
landmarks. In fact, we
the photo was taken by
rows of the sky. When th
example, DALL·E recogn
on the lights in the build

TEXT PROMPT a photo of a phone from the ...

IMAGE PROMPTS



AI-GENERATED IMAGES

We find that DALL·E has
stereotypical trends in c
over the decades. Techn
to go through periods of
dramatically shifting for
changing more increme
and streamlined.

Source

https://openai.com/blog/dall-e/

Thanks