

DALL-E^[1]

- (Dali + Wall-E)

[1] <https://openai.com/blog/dall-e/>

● Introduction

- DALL-E is a smaller version of GPT-3^[2] and it contains 12 billion parameters
- It also builds on concepts such as ImageGPT^[3]

[2] <https://arxiv.org/abs/2005.14165>

[3] <https://openai.com/blog/image-gpt/>

GPT-3

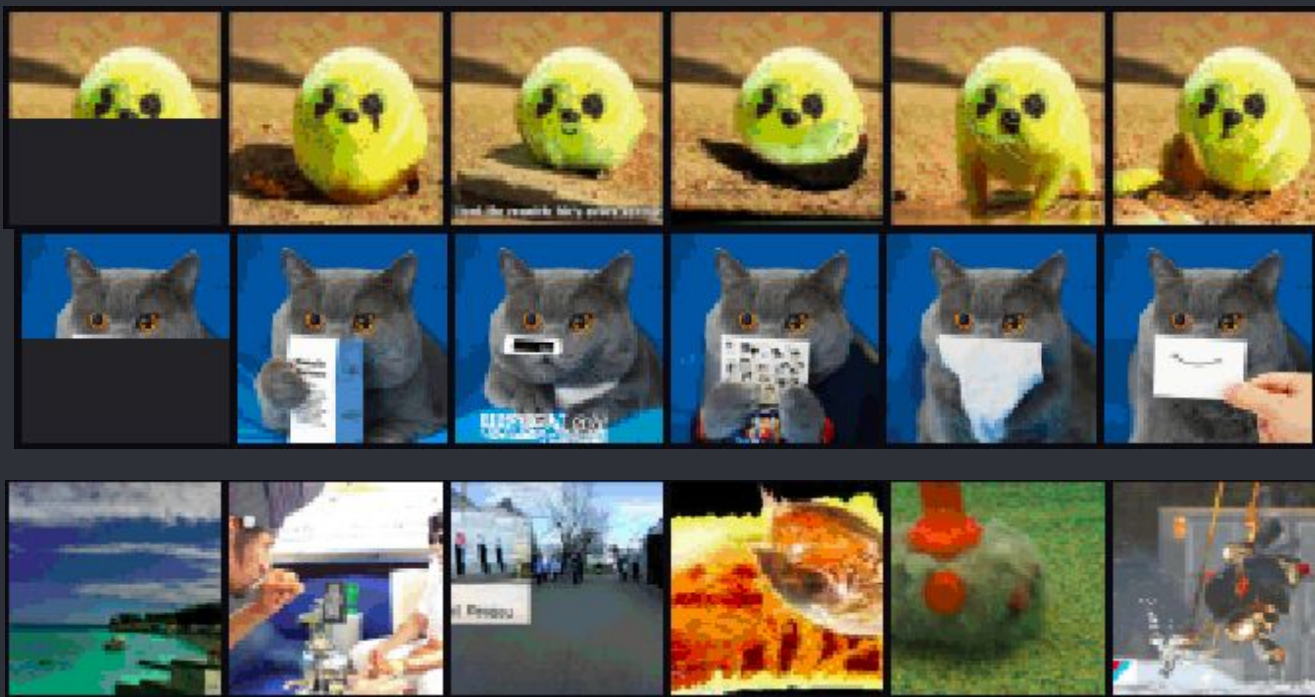
- GPT-3 is a NLP model with 175 billion parameters
 - It's an autoregressive language model
 - It is a few shot learner: it can fine-tune with a few examples and without any gradient updates on new datasets
 - Examples: It can learn to do arithmetic
 - It can learn to generate newspaper articles, given a few first sentences

	Mean accuracy	95% Confidence Interval (low, hi)	<i>t</i> compared to control (<i>p</i> -value)	"I don't know" assignments
Control (deliberately bad model)	86%	83%–90%	-	3.6 %
GPT-3 Small	76%	72%–80%	3.9 (2e-4)	4.9%
GPT-3 Medium	61%	58%–65%	10.3 (7e-21)	6.0%
GPT-3 Large	68%	64%–72%	7.3 (3e-11)	8.7%
GPT-3 XL	62%	59%–65%	10.7 (1e-19)	7.5%
GPT-3 2.7B	62%	58%–65%	10.4 (5e-19)	7.1%
GPT-3 6.7B	60%	56%–63%	11.2 (3e-21)	6.2%
GPT-3 13B	55%	52%–58%	15.3 (1e-32)	7.1%
GPT-3 175B	52%	49%–54%	16.9 (1e-34)	7.8%

Table 3.11: Human accuracy in identifying whether short (~200 word) news articles are model generated. We find that human accuracy (measured by the ratio of correct assignments to non-neutral assignments) ranges from 86% on the control model to 52% on GPT-3 175B. This table compares mean accuracy between five different models, and shows the results of a two-sample T-Test for the difference in mean accuracy between each model and the control model (an unconditional GPT-3 Small model with increased output randomness).

ImageGPT

- Based on the GPT-2 architecture
- Is trained to do next pixel prediction, so it can do image completion
- It is also possible to sample from ImageGPT



DALL-E

- DALL-E is a transformer model
- DALL-E receives both the text and the image as a single stream of data
- Images are preprocessed to 256x256, each image is then compressed to a 32x32 grid of discrete latent codes using a pre-trained discrete VAE.
- The way it maps from text to images is using 'tokens':
 - Each letter in the alphabet could be a token: 26
 - Each image caption is represented using a maximum of 256 BPE tokens
 - BPE = Byte pair encoding
 - The vocabulary size for the captions is 16384
 - The image is represented using 1024 tokens (32x32) and has a vocabulary size of 8192 (8*1024)

DALL-E

- DALL-E is a transformer model
- DALL-E receives both the text and the image as a single stream of data
- Images are preprocessed to 256x256, each image is then compressed to a 32x32 grid of discrete latent codes using a pre-trained discrete VAE.
- The way it maps from text to images is using 'tokens':
 - Each letter in the alphabet could be a token: 26
 - Each image caption is represented using a maximum of 256 BPE tokens
 - BPE = Byte pair encoding
 - The vocabulary size for the captions is 16384
 - The image is represented using 1024 tokens (32x32) and has a vocabulary size of 8192 (8*1024)

● DALL-E

- The model can draw multiple objects, their attributes, and their spatial relationship
 - Depends on how the caption is phrased
- I'll go through some examples on the website

● DALL-E

- Implementation:
 - It models the tokens autoregressively
 - It has 64 self-attention layers and it uses an attention mask to allow each image token to attend to all caption tokens
 - → Full method will be published in upcoming paper

Thanks!

ANY QUESTIONS?

Let's keep discussing the ideas and looking for ways to learn them deeper by applying them in unexpected ways

1

TRANSITION HEADLINE

Let's start with the first set of slides

“

Quotations are commonly printed as a means of inspiration and to invoke philosophical thoughts from the reader.

● THIS IS A SLIDE TITLE

○ Here you have:

- A list of items
- And some text
- But remember not to overload your slides with content

Your audience will listen to you or read the content, but won't do both.



BIG CONCEPT

Bring the attention of your audience over a key concept using icons or illustrations

● YOU CAN ALSO SPLIT YOUR CONTENT

○ **White**

Is the color of milk and fresh snow, the color produced by the combination of all the colors of the visible spectrum.

Black

Is the color of coal, ebony, and of outer space. It is the darkest color, the result of the absence of or complete absorption of light.

IN TWO OR THREE COLUMNS

Yellow

Is the color of gold, butter and ripe lemons. In the spectrum of visible light, yellow is found between green and orange.

Blue

Is the colour of the clear sky and the deep sea. It is located between violet and green on the optical spectrum.

Red

Is the color of blood, and because of this it has historically been associated with sacrifice, danger and courage.

- A PICTURE IS WORTH A THOUSAND WORDS



A complex idea can be conveyed with just a single still image, namely making it possible to absorb large amounts of data quickly.

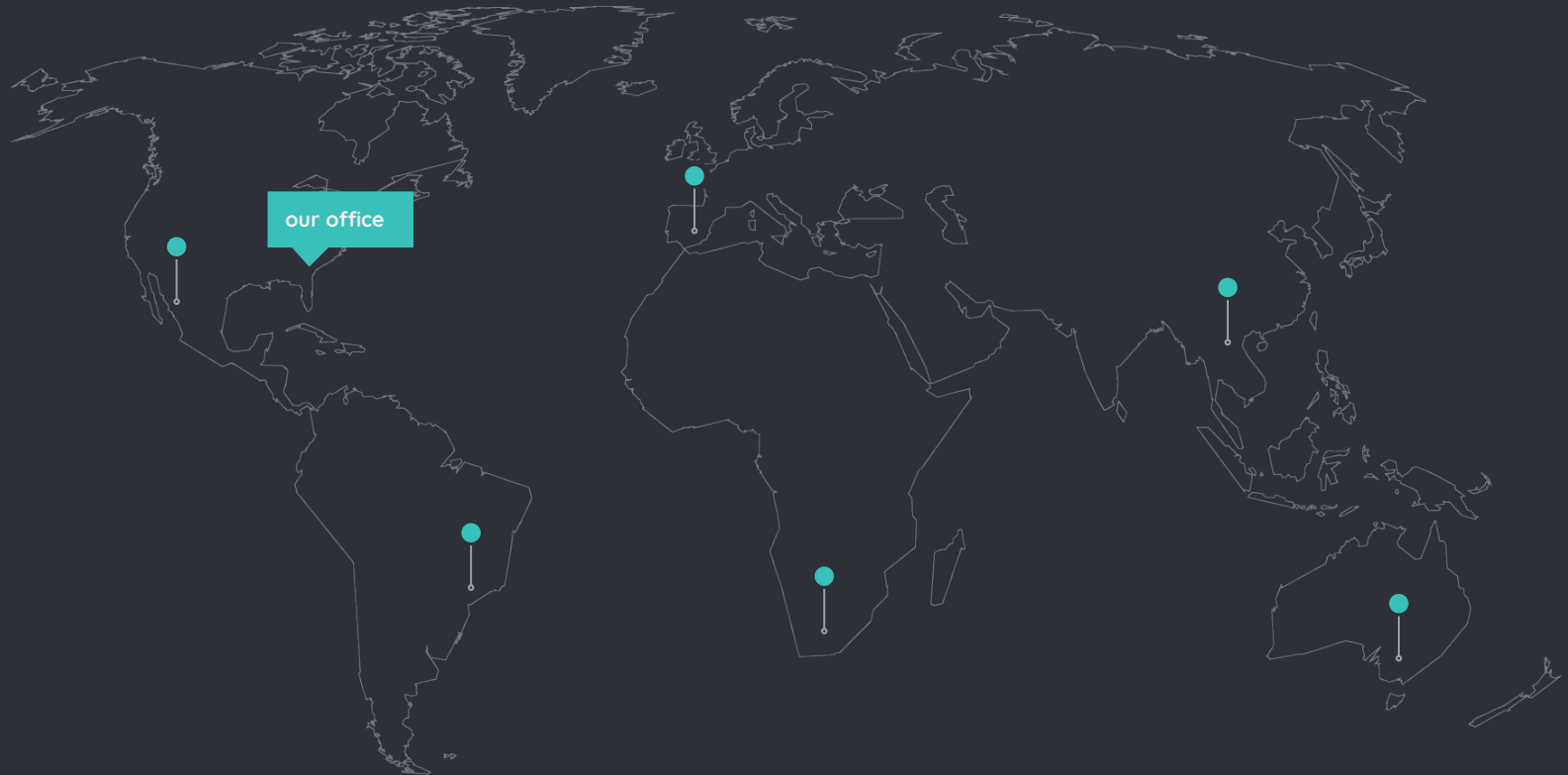
A close-up photograph of a person's hand holding a smartphone. The phone's screen displays a high-angle, wide-area photograph of a city, likely New York City, showing a dense urban landscape with numerous skyscrapers and a large body of water in the background. The image on the screen is vibrant and detailed. On the right side of the phone's screen, there are status icons: a battery level indicator at the top, followed by 'WDR OFF' and a camera icon at the bottom. The hand holding the phone has dark nail polish. To the left of the phone, there is a teal-colored graphic element consisting of a small circle and a vertical line.

WANT BIG IMPACT?
Use big image.

● AND TABLES TO COMPARE DATA

	A	B	C
Yellow	10	20	7
Blue	30	15	10
Orange	5	24	16

MAPS





89,526,124

Whoa! That's a big number,
aren't you proud?



89,526,124\$

That's a lot of money

185,244 users

And a lot of users

100%

Total success!

LET'S REVIEW SOME CONCEPTS



Yellow

Is the color of gold, butter and ripe lemons. In the spectrum of visible light, yellow is found between green and orange.



Blue

Is the colour of the clear sky and the deep sea. It is located between violet and green on the optical spectrum.



Red

Is the color of blood, and because of this it has historically been associated with sacrifice, danger and courage.



Yellow

Is the color of gold, butter and ripe lemons. In the spectrum of visible light, yellow is found between green and orange.



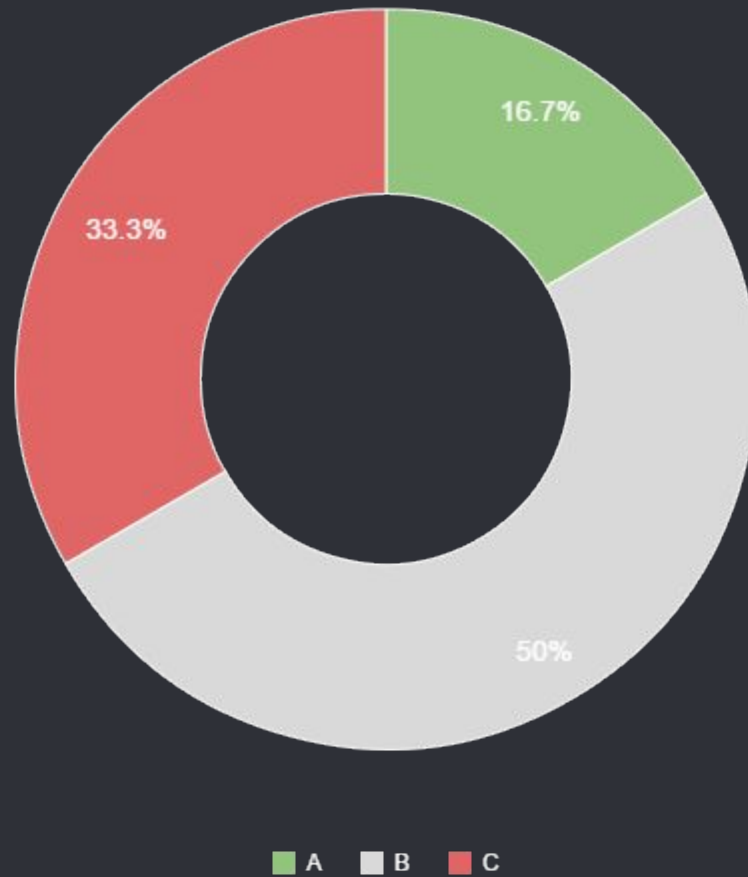
Blue

Is the colour of the clear sky and the deep sea. It is located between violet and green on the optical spectrum.



Red

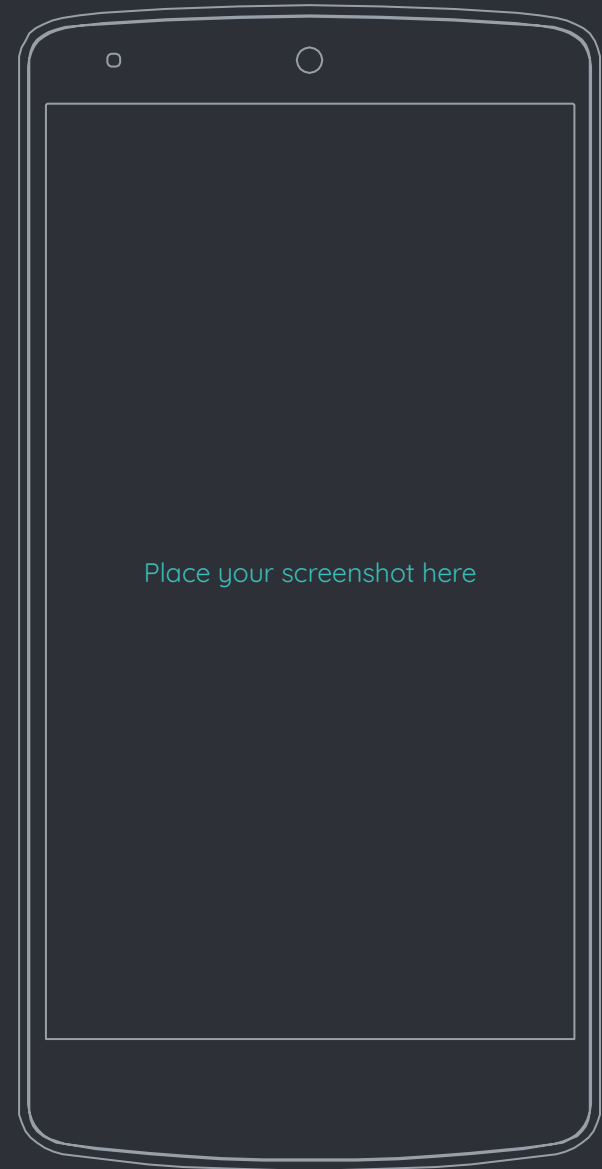
Is the color of blood, and because of this it has historically been associated with sacrifice, danger and courage.



○ You can copy&paste graphs from Google Sheets

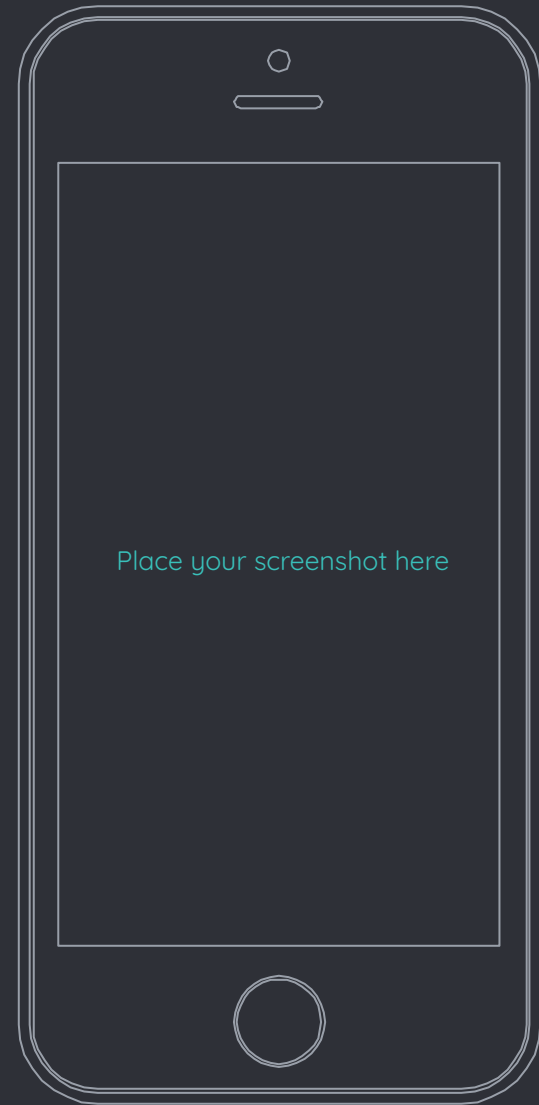
ANDROID PROJECT

Show and explain your web, app or software projects using these gadget templates.



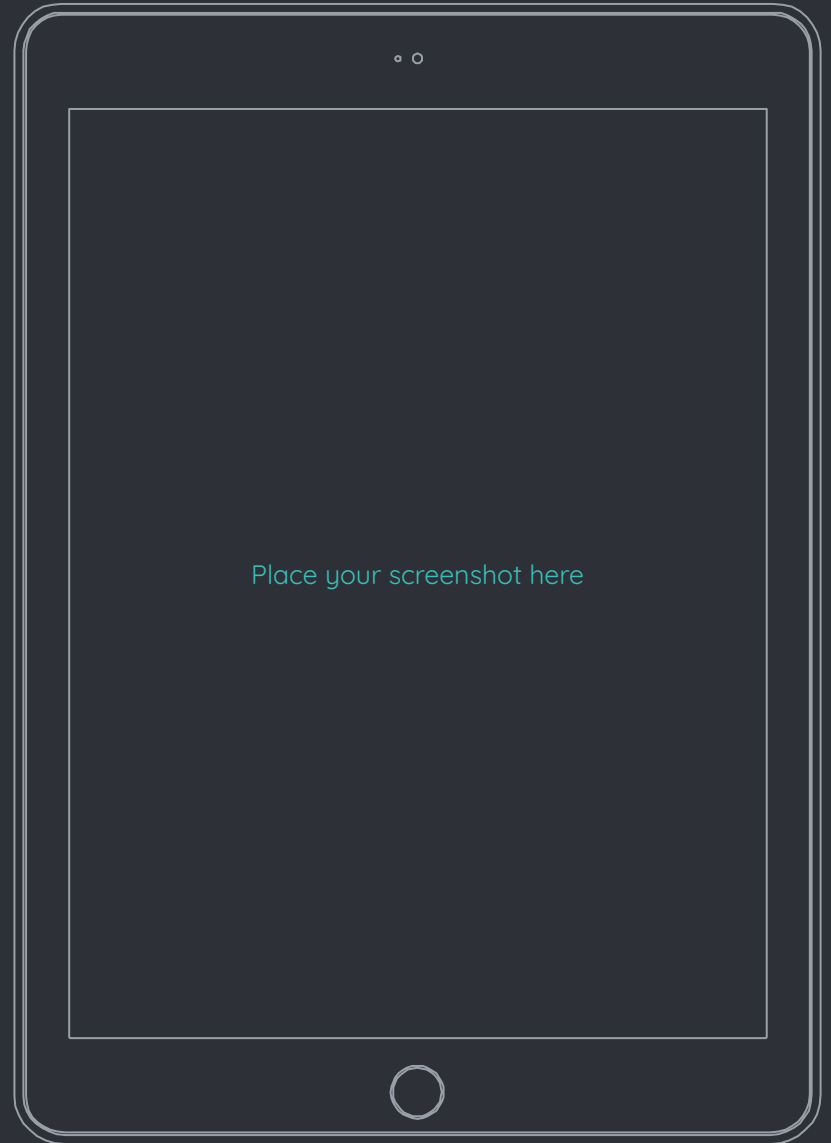
IPHONE PROJECT

Show and explain your web, app or software projects using these gadget templates.



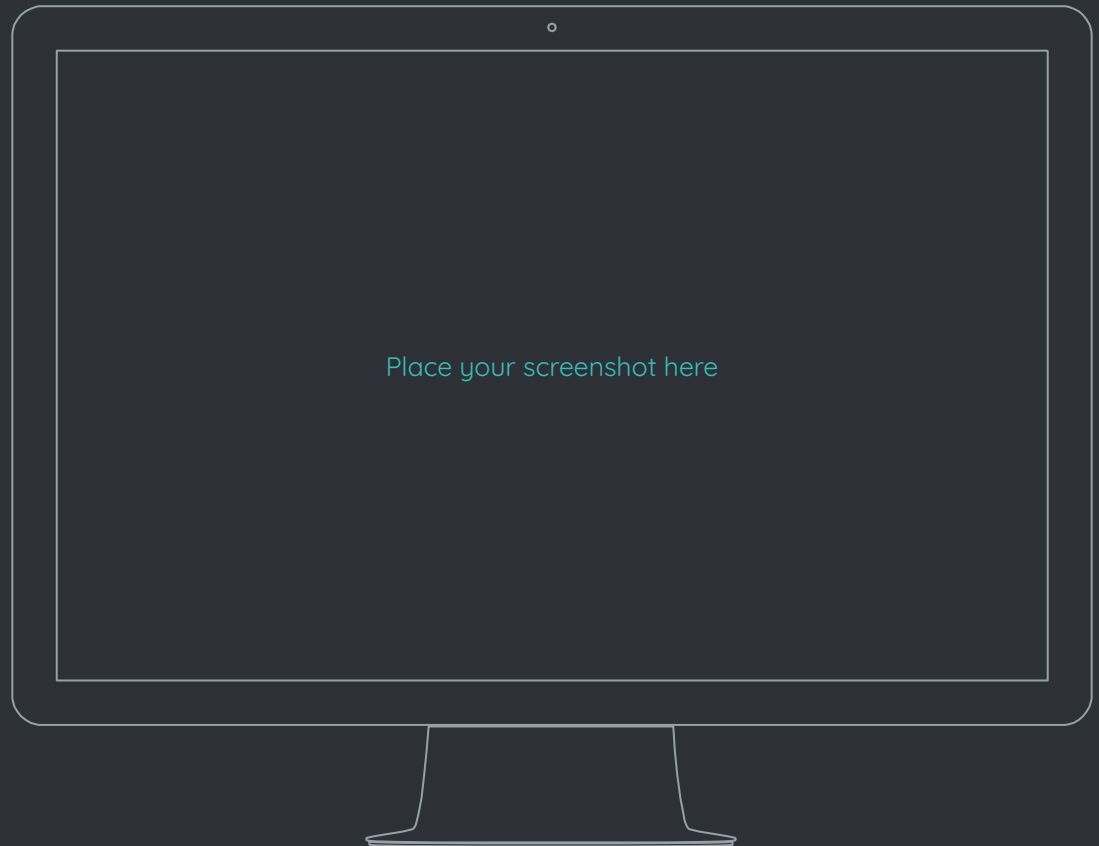
TABLET PROJECT

Show and explain your web, app or software projects using these gadget templates.



DESKTOP PROJECT

Show and explain your web, app or software projects using these gadget templates.



Thanks!

ANY QUESTIONS?

You can find me at

@username

user@mail.me

● CREDITS

○ Special thanks to all the people who made and released these awesome resources for free:

- Presentation template by SlidesCarnival
- Photographs by Unsplash

PRESENTATION DESIGN

This presentations uses the following typographies and colors:

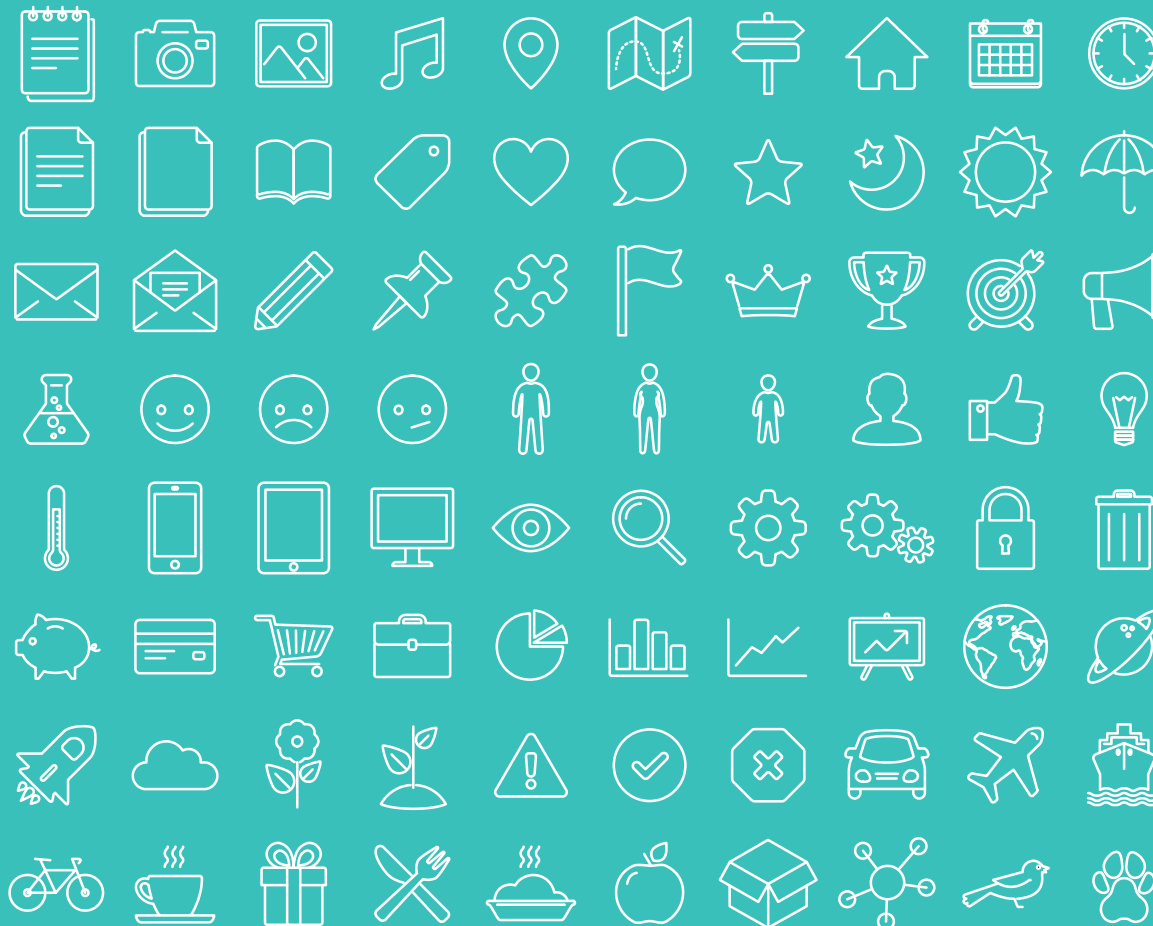
- Titles & body copy: **Quicksand**

You can download the fonts on this page:

<https://www.fontsquirrel.com/fonts/quicksand>

- Dark gray **#2e3037**
- Aqua **#39c0ba**
- Salmon **#f35b69**
- Blue **#6d9eeb**

You don't need to keep this slide in your presentation. It's only here to serve you as a design guide if you need to create new slides or download the fonts to edit the presentation in PowerPoint®



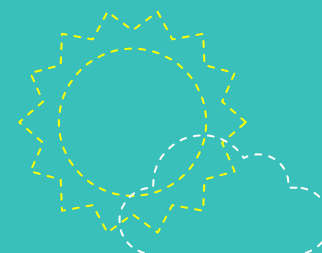
**SlidesCarnival icons are
editable shapes.**

This means that you can:

- Resize them without losing quality.
- Change line color, width and style.

Isn't that nice? :)

Examples:



Now you can use any emoji as an icon!

And of course it resizes without losing quality and you can change the color.

How? Follow Google instructions

<https://twitter.com/googledocs/status/730087240156643328>



and many more...