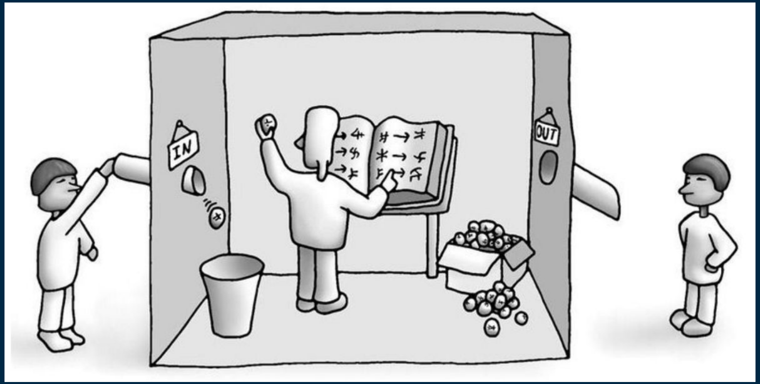
SELF SUPERVISED LEARNING

SimCLR & NLP

By
Milis Sahar

The Chinese Room Argument

First published Fri Mar 19, 2004; substantive revision Thu Feb 20, 2020



"It has become one of the best-known arguments in recent philosophy."

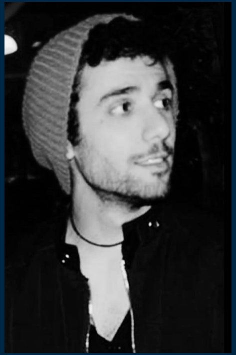
“Self-supervised learning is the dark matter of intelligence”

Yan LaCun, FAIR

(DM is 85% of the matter in the universe)

Turn left!

ABOUT ME



SAHAR MILLIS

NLP DATA SCIENTIST @VIMEO

Student & T.A, MSc Data Science @RU

https://medium.com/@sahar.millis

ON THE NEXT 30 MINUTES...

01

SLL

Timeline, Data, Methods

02

SimCLR

Article, Training, Findings

03

SLL & NLP

LMs, SimCSE

SSL

Self-Supervised Learning is an ML method.
It learns from unlabeled sample data.
It's an intermediate form between supervised and unsupervised learning.

Self-supervised learning has produced promising results in recent years.

The primary appeal of SSL is that training can occur with data of lower quality.

Self-supervised learning more closely imitates the way humans learn.

SSL TIMELINE (models)

2018

Unsupervised Translation
Meta

2019

RoBERTa
Wave2Vec

2020

ViLBERT, RIO
SimSearch

2021

Textless NLP
PyTorchVideo
HuBERT

You’re using SSL every day…

WHERE IS SSL?

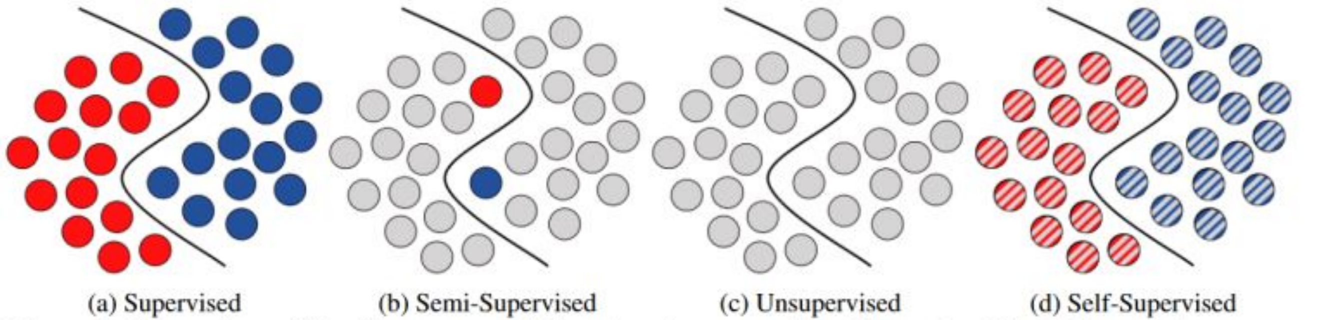
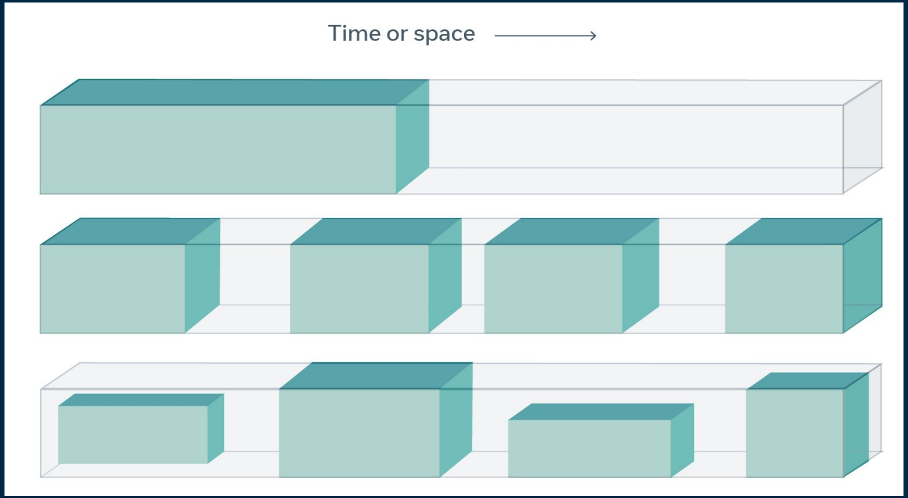


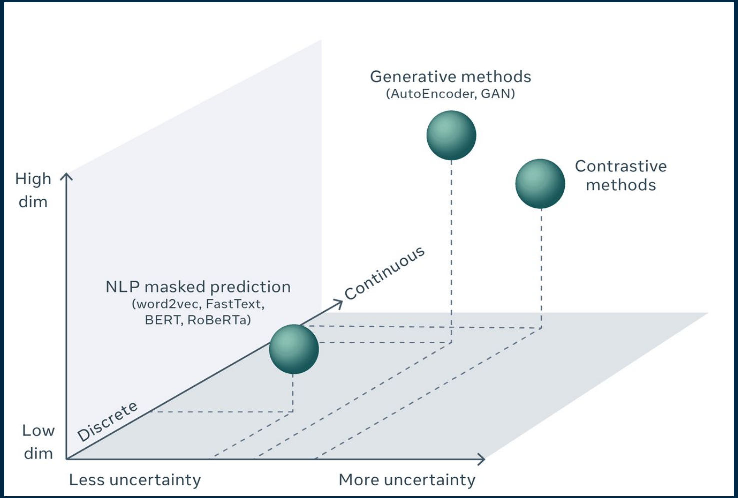
Figure 2: Illustrations of the four presented deep learning strategies - The red and dark blue circles represent labeled data points of different classes. The light grey circles represent unlabeled data points. The black lines define the underlying decision boundary between the classes. The striped circles represent datapoints which ignore and use the label information at different stages of the training process.

Semi-, Self- and Unsupervised Learning in Image Classification (May, 2021)

DATA

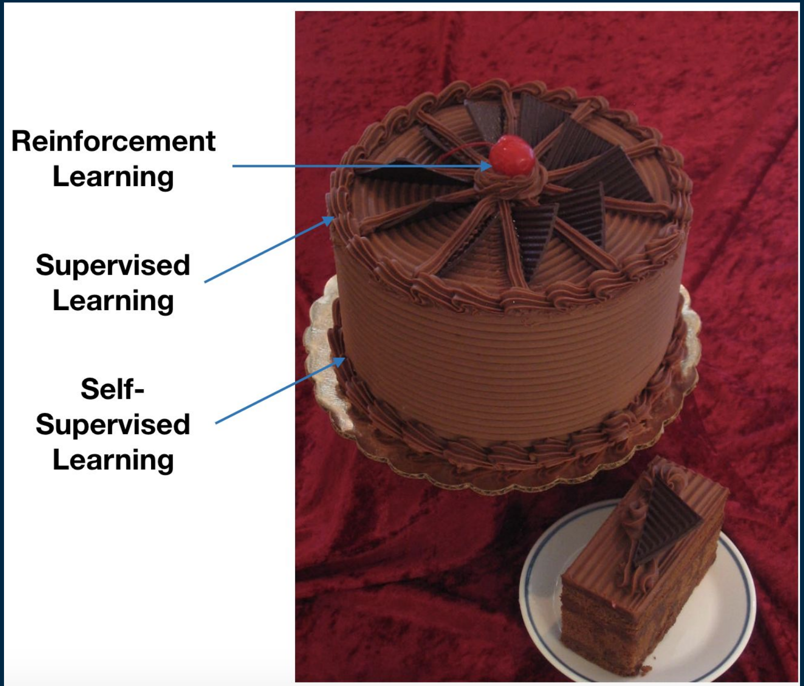
Self-supervised learning obtains supervisory signals from the data itself, often leveraging the underlying structure in the data.



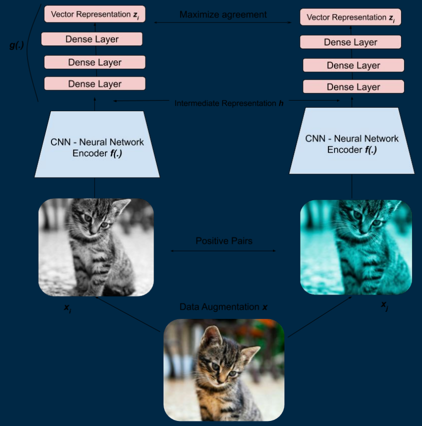


PREDICTIVE LEARNING

"cake analogy"
by Prof. Yan LaCun
At the NIPS 2016 conference
Referring to Predictive Learning

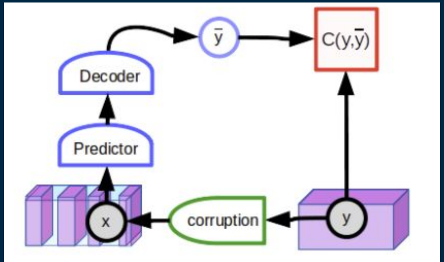


SSL TECHNIQUES



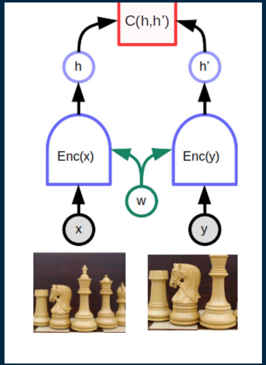
constructing pairs of x and y that are not compatible

Contrastive



measures distance between vectors produced by identical networks

Siamese networks



Non-contrastive

Deeper-Cluster, ClusterFit, MoCo-v2, SwAV, SimSiam

Contrastive vs Non-Contastive

	CONTRASTIVE	NON-CONTRASTIVE
Samples	Negative & positive	Positive
Approach	Minimise distance for positive Maximise distance for negative	Leverages positive data
Network	More Complex	Simpler

SSL TASKS

ervation x and a proposed prediction y.

Joint embedding

Segmentation, Categories, Semantic Analysis, etc.

Non-contrastive

Generate Text2Text, Text2Code, Image2Text, etc.

Siamese networks

Classify unstructured text into names, organizations, locations, dates, quantities, etc.

Named Entity Recognition

constructing pairs of x and y that are not compatible,

Contrastive

Semantic Similarity, Relation Predictions, etc.

Text Pair Matching

SimCLR Jul 2020

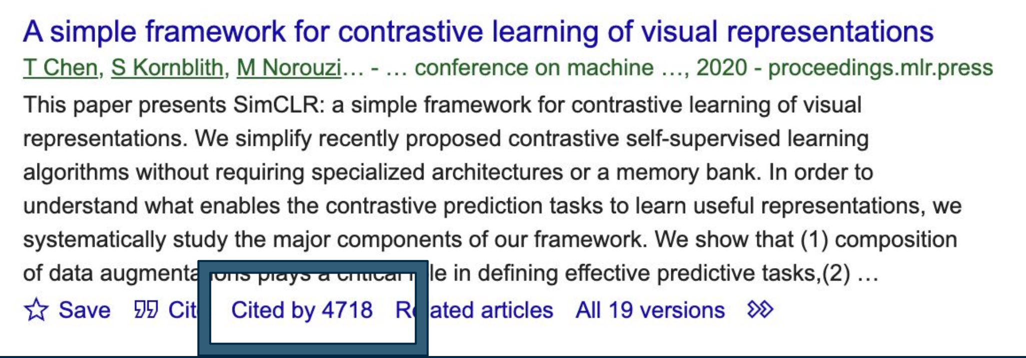
A Simple Framework for Contrastive Learning of Visual Representations

SimCLR Article

Almost 5K cited

Using Contrastive Loss as SSL

Key findings: Augmentation, Projection, Scale



Contrastive Loss

Key Ideas:

Same Class -> Similar Embeddings

Different Class -> Dissimilar Embeddings

(Challenge: How do you sample?)

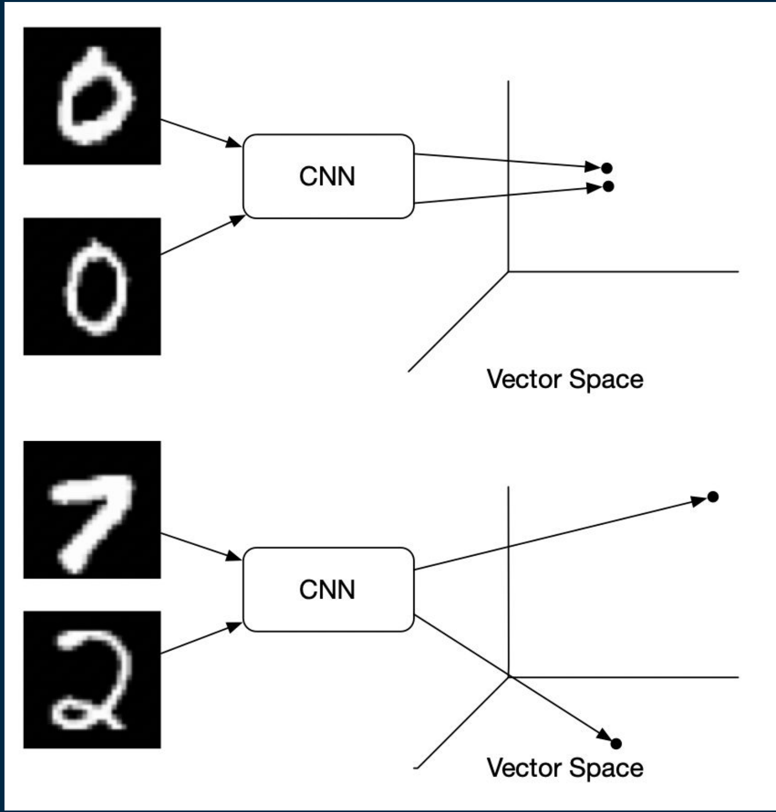
For positive pair (i,j) :

$$softmax(Z_i) = \frac{\exp(Z_i)}{\sum \exp(Z_i)}$$

Similarity of positive pair

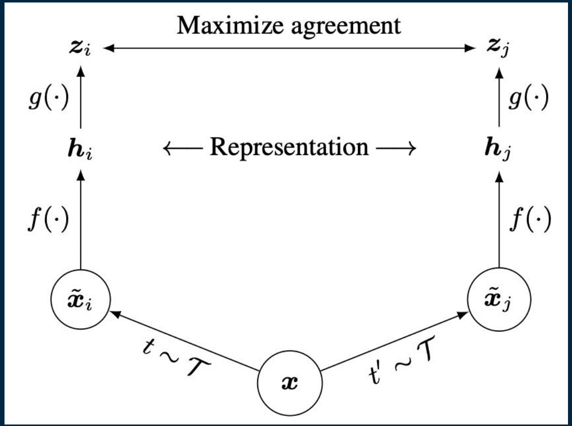
$$\ell_{i,j} = -\log \frac{\exp(\boxed{\mathrm{sim}(\boldsymbol{z}_i, \boldsymbol{z}_j)}/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\boxed{\mathrm{sim}(\boldsymbol{z}_i, \boldsymbol{z}_k)}/\tau)}$$

Similarity of negatives



Key Idea: Augmentation

Generate augmentations for positive and negative samples



Batch of N images:

Two augs of same image

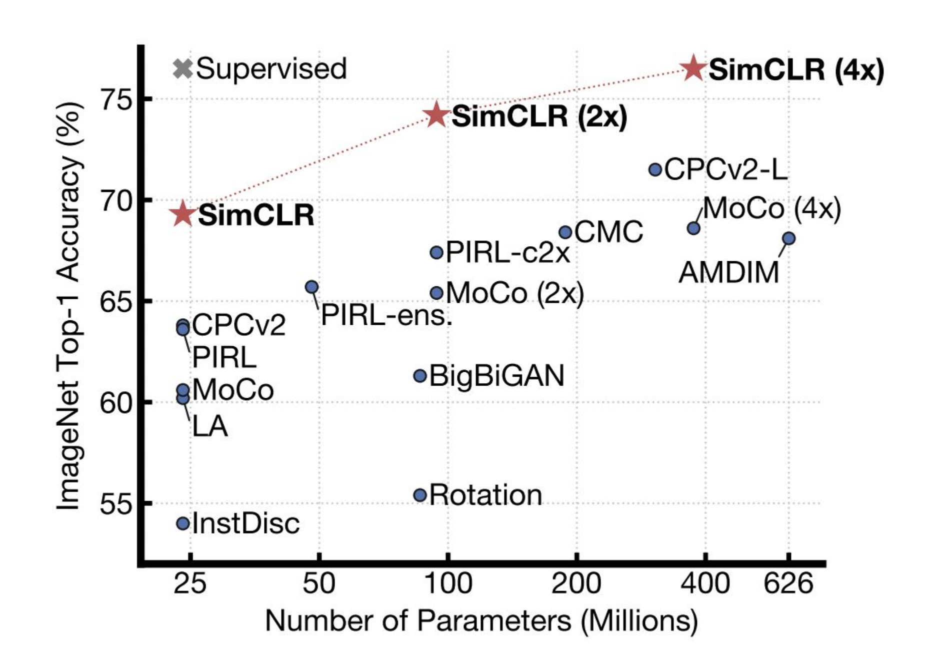
$$\ell_{i,j} = -\log \frac{\exp(\boxed{\mathrm{sim}(\boldsymbol{z}_i, \boldsymbol{z}_j)}/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\boxed{\mathrm{sim}(\boldsymbol{z}_i, \boldsymbol{z}_k)}/\tau)}$$

All the other images in the batch

Evaluation

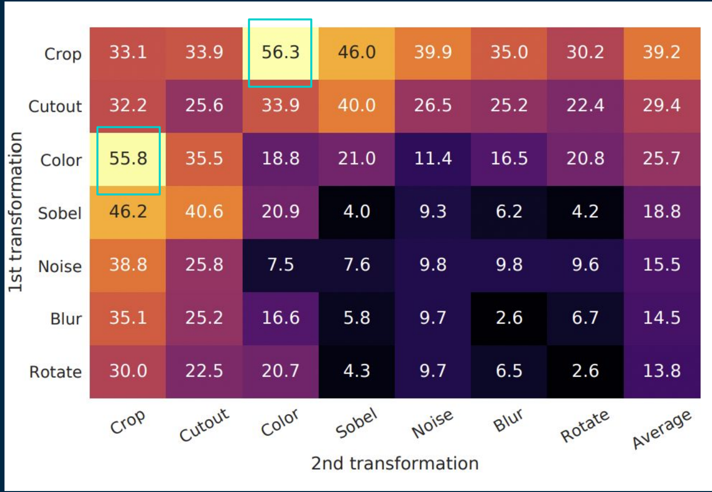
Strong SS performance

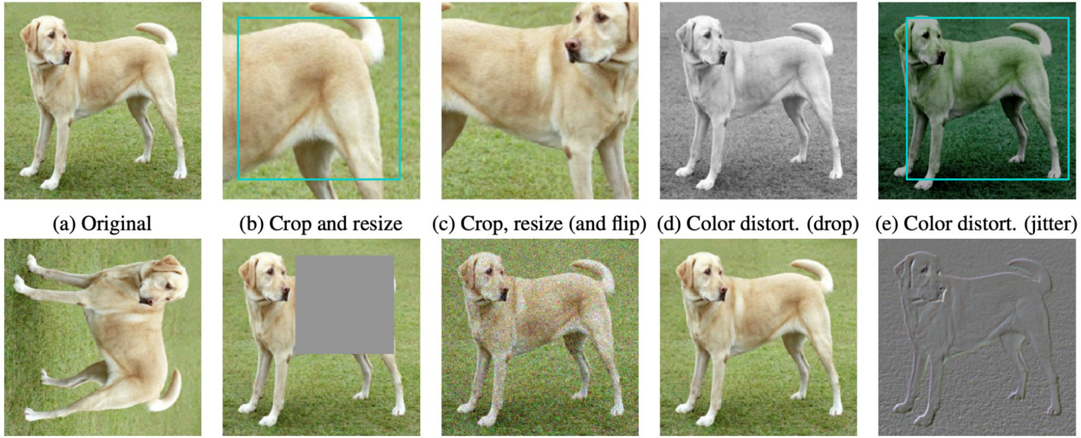
Method	Architecture	Label fraction 1%	10%
		Top 5	
Supervised baseline	ResNet-50	48.4	80.4
Methods using other label-propagation:			
Pseudo-label	ResNet-50	51.6	82.4
VAT+Entropy Min.	ResNet-50	47.0	83.4
UDA (w. RandAug)	ResNet-50	-	88.5
FixMatch (w. RandAug)	ResNet-50	-	89.1
S4L (Rot+VAT+En. M.)	ResNet-50 (4×)	-	91.2
Methods using representation learning only:			
InstDisc	ResNet-50	39.2	77.4
BigBiGAN	RevNet-50 (4×)	55.2	78.8
PIRL	ResNet-50	57.2	83.8
CPC v2	ResNet-161(∗)	77.9	91.2
SimCLR (ours)	ResNet-50	75.5	87.8
SimCLR (ours)	ResNet-50 (2×)	83.0	91.2
SimCLR (ours)	ResNet-50 (4×)	**85.8**	**92.6**



Key Findings – Augmentation

Which augmentations matter?





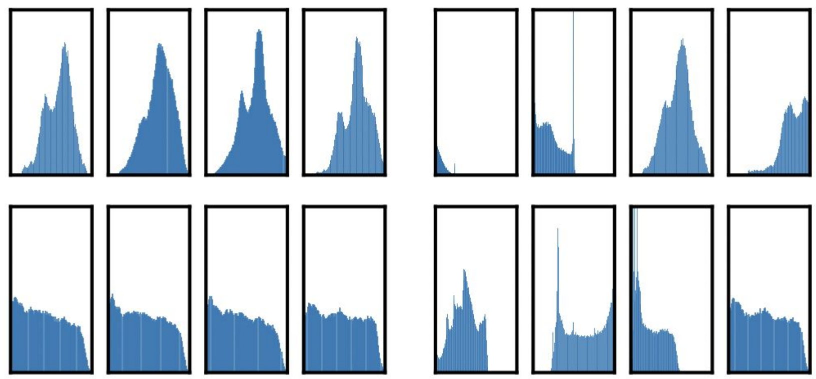
(a) Original (b) Crop and resize (c) Crop, resize (and flip) (d) Color distort. (drop) (e) Color distort. (jitter)

(f) Rotate $\{90°, 180°, 270°\}$ (g) Cutout (h) Gaussian noise (i) Gaussian blur (j) Sobel filtering

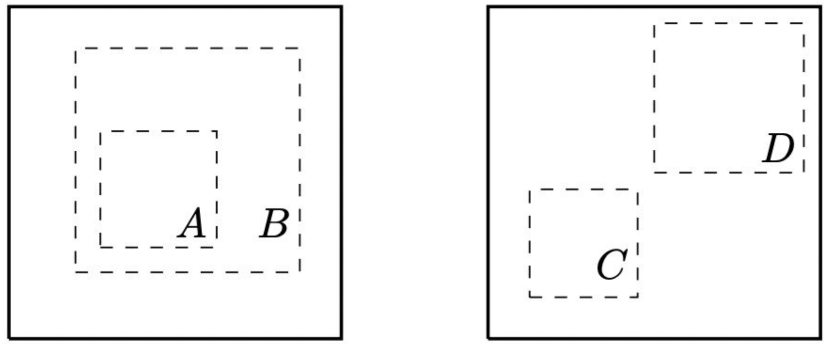
Key Findings – Augmentation

Why cropping?

Why and When to use color distortion?



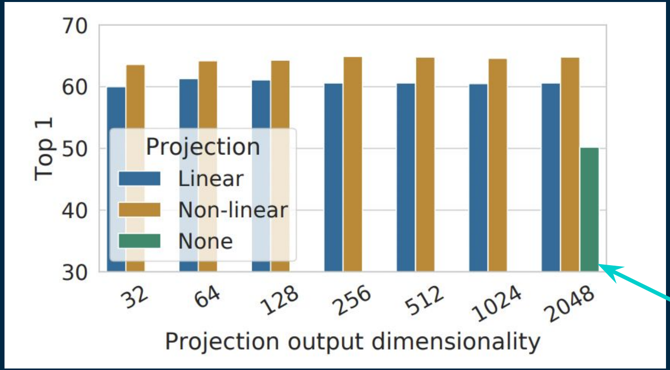
(a) Without color distortion. (b) With color distortion.



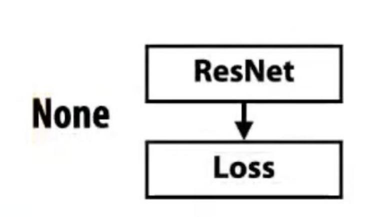
(a) Global and local views. (b) Adjacent views.

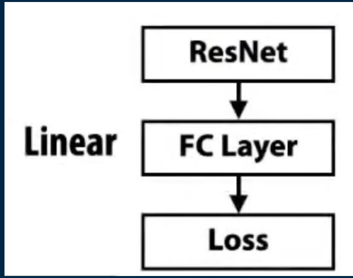
Key Findings – Projection

None | Linear | Non-Linear

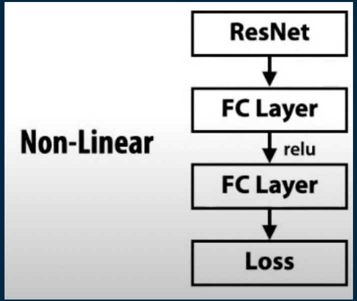


Not an error





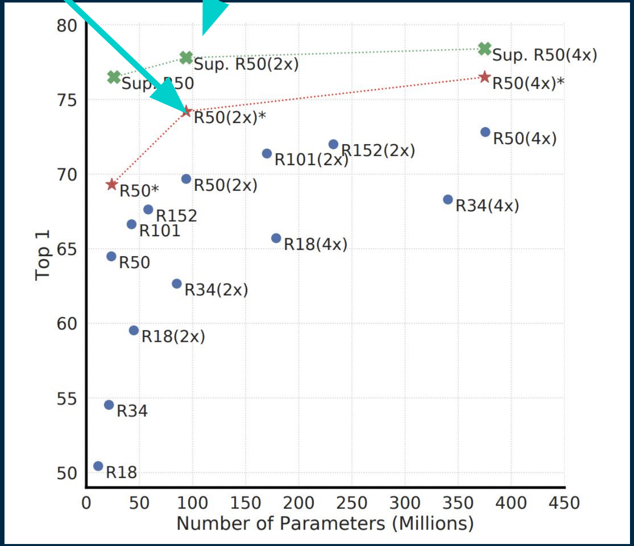
USE THIS ONE!



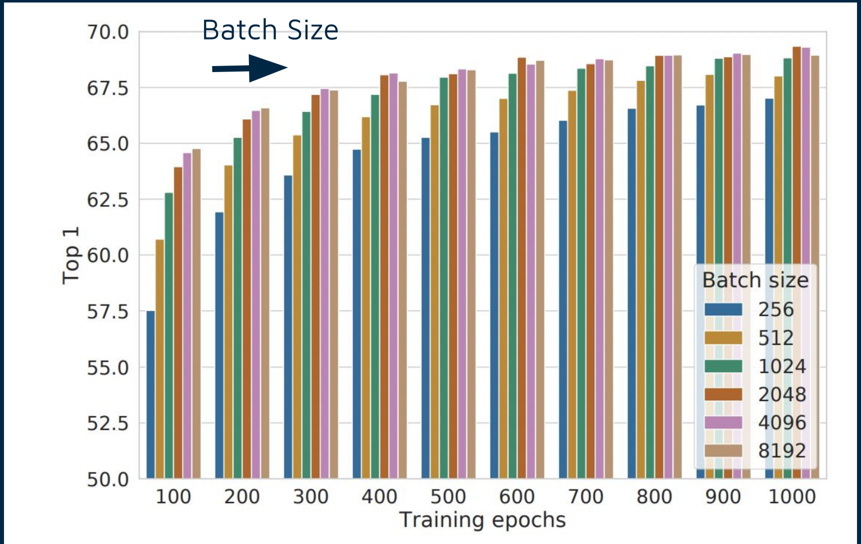
Key Findings – Scale

SimSLR

ResNet



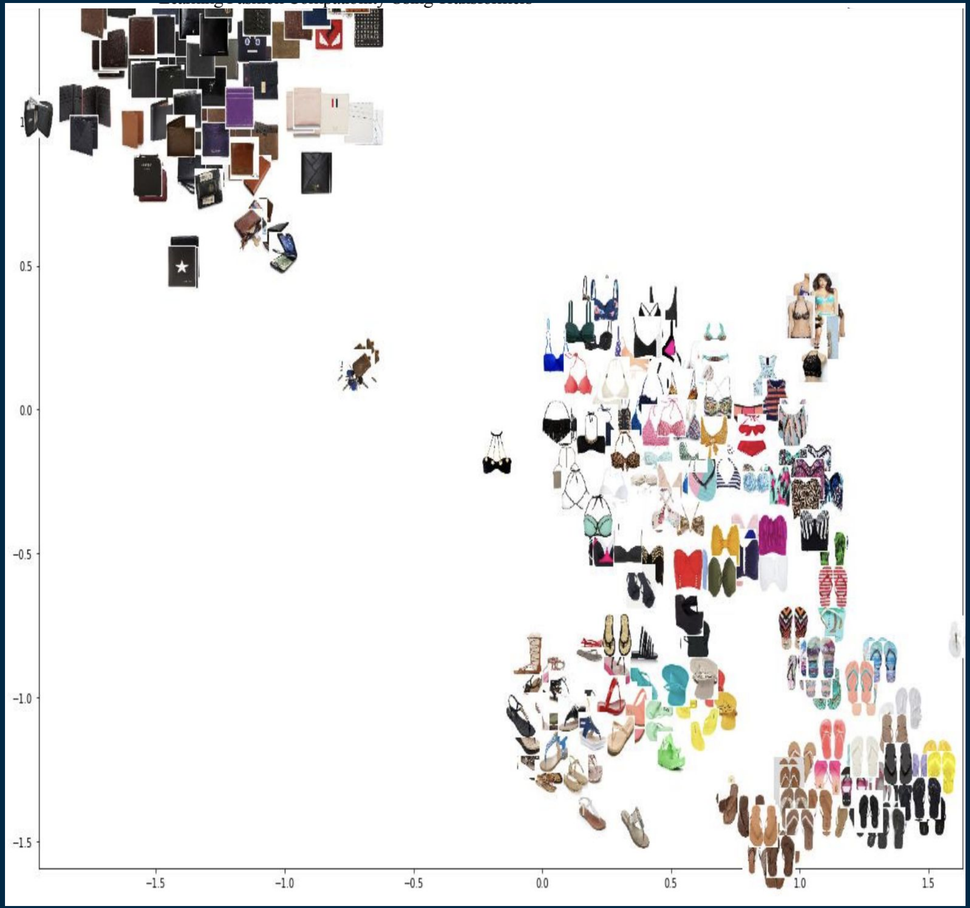
Model Size



Training time

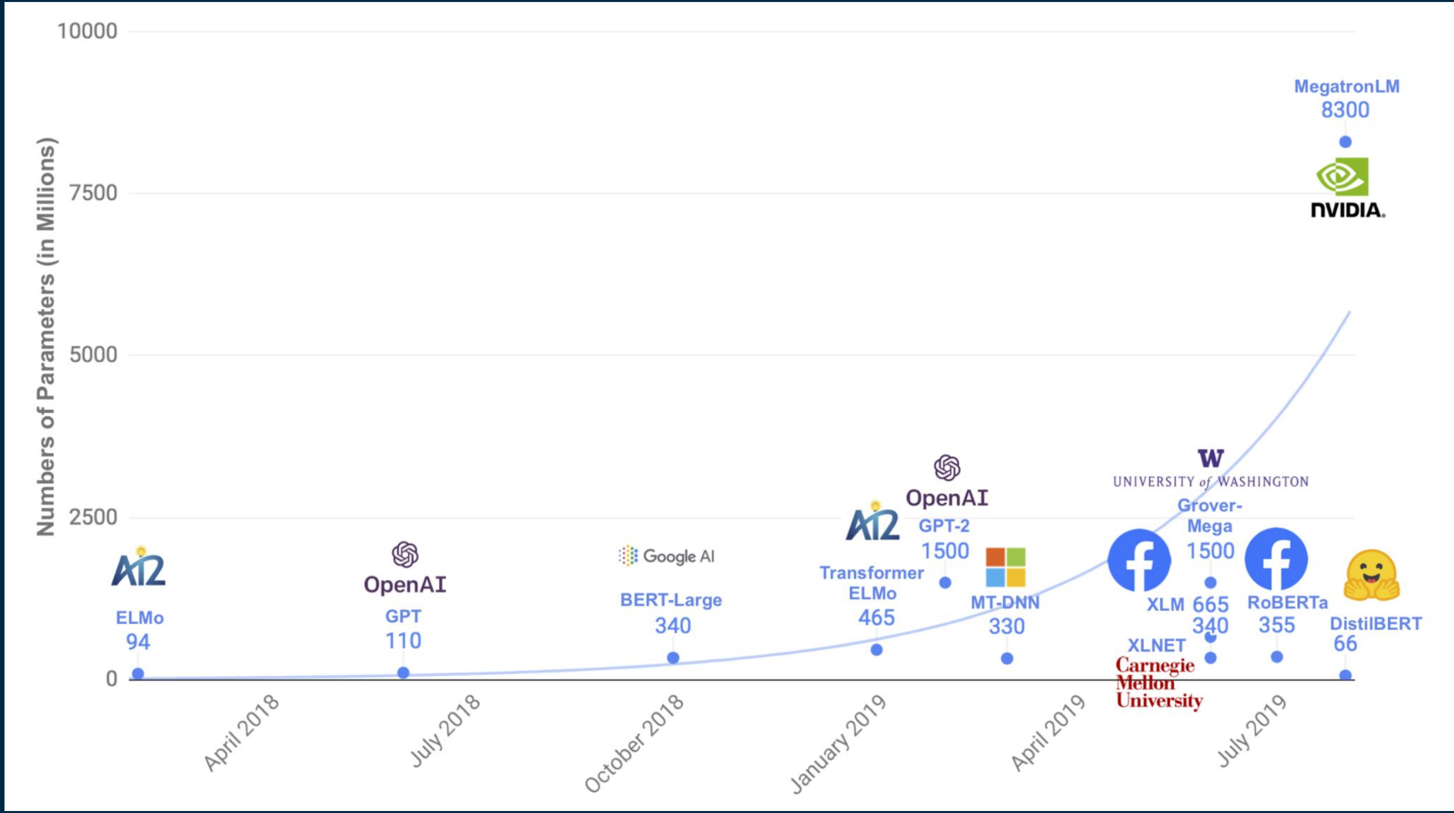
SimCLR on different dataset

Using SimCLR on a unique fashion dataset



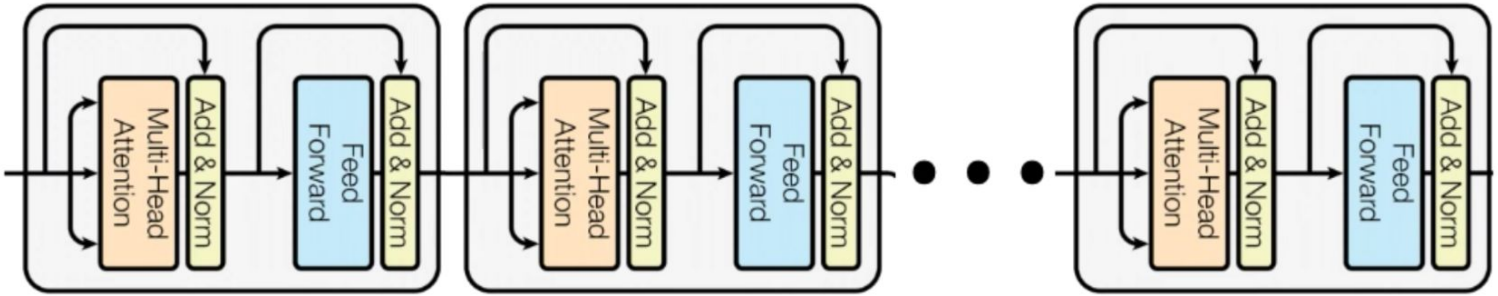
SLL with LANGUAGE MODELS

From the article: DistilBERT

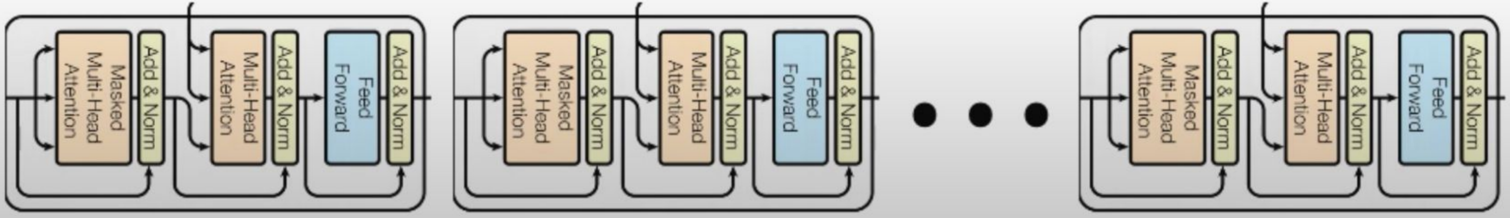


8,300,000,000

MEGATRON-LM


Multi-Head Attention
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Add & Norm
Feed Forward
Add & Norm

BERT


Masked Multi-Head Attention
Add & Norm
Multi-Head Attention
Add & Norm
Feed Forward
Add & Norm
Masked Multi-Head Attention
Add & Norm
Multi-Head Attention
Add & Norm
Feed Forward
Add & Norm
Masked Multi-Head Attention
Add & Norm
Multi-Head Attention
Add & Norm
Feed Forward
Add & Norm

GPT

BERT (Google, October 2018)

Bidirectional Encoder Representations from Transformers

PRETRAIN

Mask Language Modeling (MLM)

Bank of <mask> river.

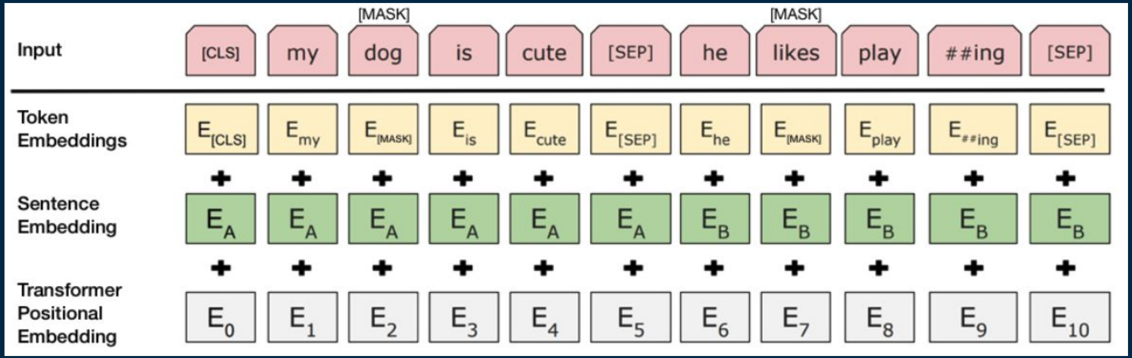
Next Sentence Prediction (NSP)

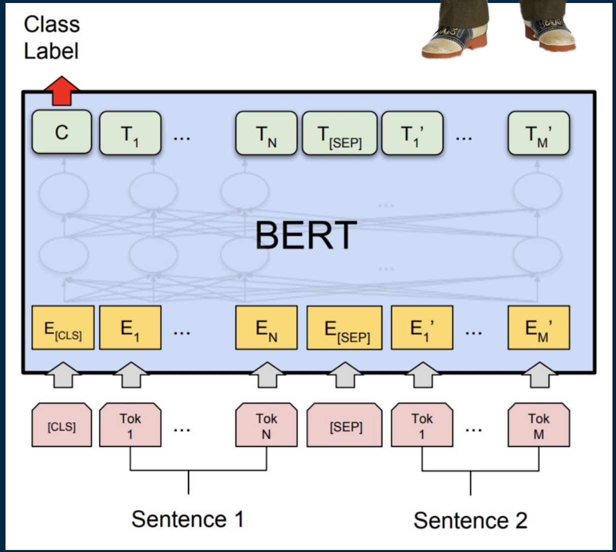
Data Science is cool. I like it.

FINE TUNE

Add FC layer on top of BERT

Train on Classification, Q&A, etc.





GPT-3 (OpenAI, May 2020)

Language Models are Few-Shot Learners

Zero-Shot
The model predicts the answer given only a natural language description.

No gradient updates are performed.

`Translate English to French:`	*task description*
`cheese =>`	*prompt*

One-Shot
In addition to the task description, the model sees a single example of the task.

No gradient updates are performed.

`Translate English to French:`	*task description*
`sea otter => loutre de mer`	*example*
`cheese =>`	*prompt*

Few-Shot
In addition to the task description, the model sees a few examples of the task.

No gradient updates are performed.

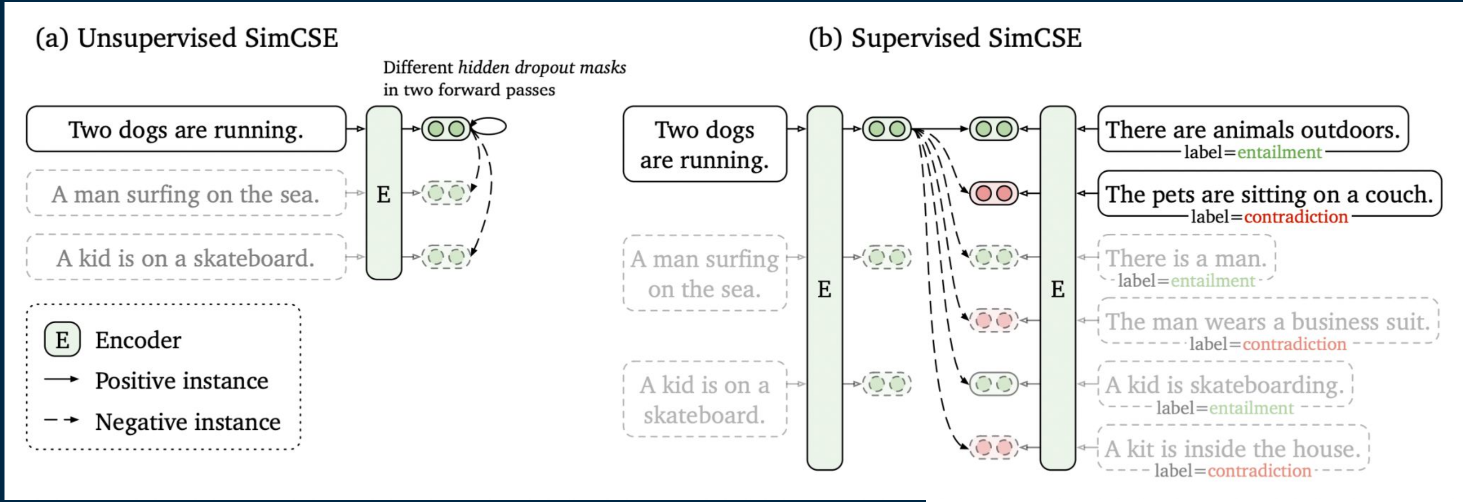
`Translate English to French:`	*task description*
`sea otter => loutre de mer`	*examples*
`peppermint => menthe poivrée`	
`plush girafe => girafe peluche`	
`cheese =>`	*prompt*

SimCSE (May 2022)

Simple Contrastive Learning of Sentence Embeddings

Augmentation - Dropout

Augmentation
Synonyms, Vocabulary, Translation, etc.



Similarity funciton | Sentence embedding | Temperature

$$\ell_i = -\log \frac{e^{\mathrm{sim}(\mathbf{h}_i, \mathbf{h}_i^+)/\tau}}{\sum_{j=1}^{N} e^{\mathrm{sim}(\mathbf{h}_i, \mathbf{h}_j^+)/\tau}}$$

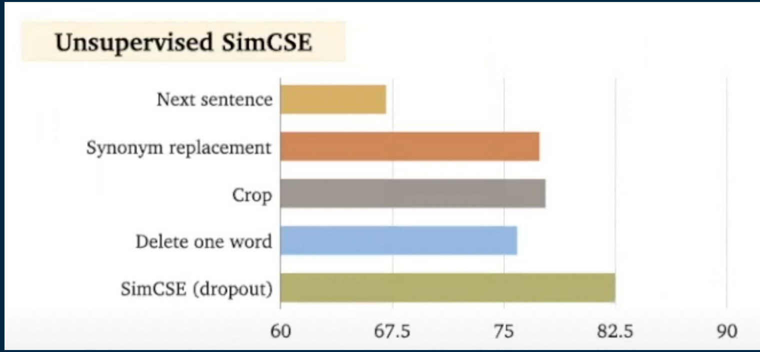
Positive pairs

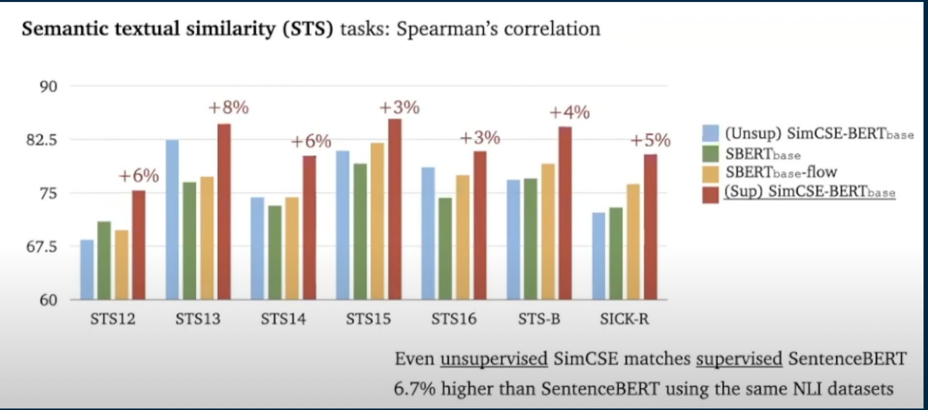
Negative pairs

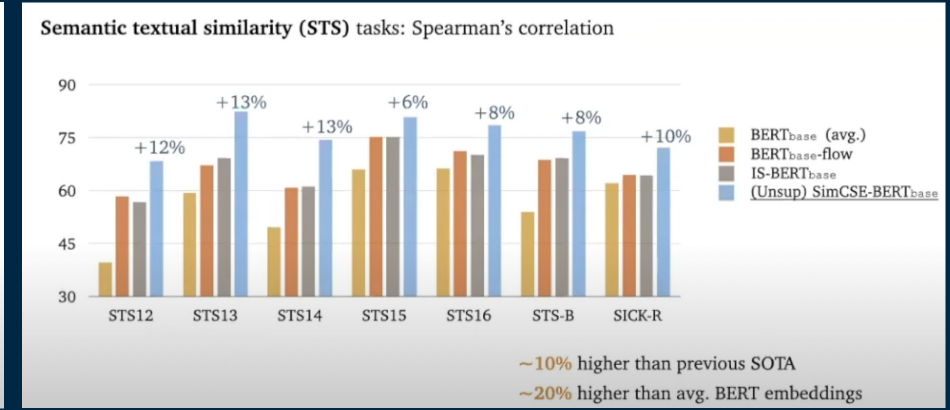
SimCSE (May 2022)

Simple Contrastive Learning of Sentence Embeddings

- Contrastive Learning improve pretrained Embedding
- Use Alignment and Uniformity to analyze different pairs







RECAP

SSL

WHAT IS SSL?

Learning from data

SimCLR

Article Review

Training, Evaluation, Findings

NLP

SLL in LANGUAGE MODELs

BERT, GPT-X, SimCSE, etc.

Do you have any questions?

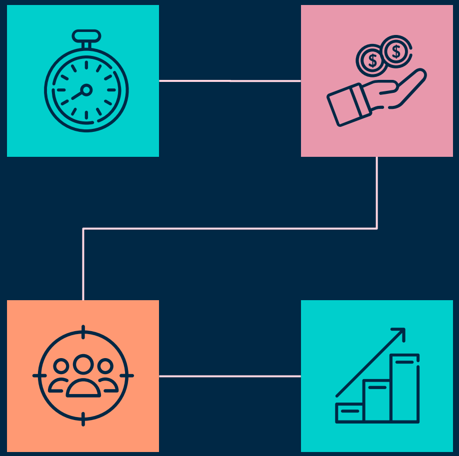
SAHAR.MILIS@google.com

https://github.com/saharmilis/
https://medium.com/@sahar.millis

THANKS

BERT – Bidirectional Encoder Representations from Transformers

NLP
PROCESS THE DATA



NLU
UNDERSTAND LANGUAGE

NLG
It’s the biggest planet in the Solar System

TAKS
Let’s talk about it :)

UNDERSTANDING THE PROBLEM

MARS

Despite being red, Mars is a cold place. It's full of iron oxide dust, which gives the planet its reddish cast

VENUS

Venus has a beautiful name and is the second planet from the Sun. It's terribly hot, even hotter than Mercury

MAIN COMPETITORS

NEPTUNE

It’s the farthest planet from the Sun

MARS

Despite being red, Mars is a cold place

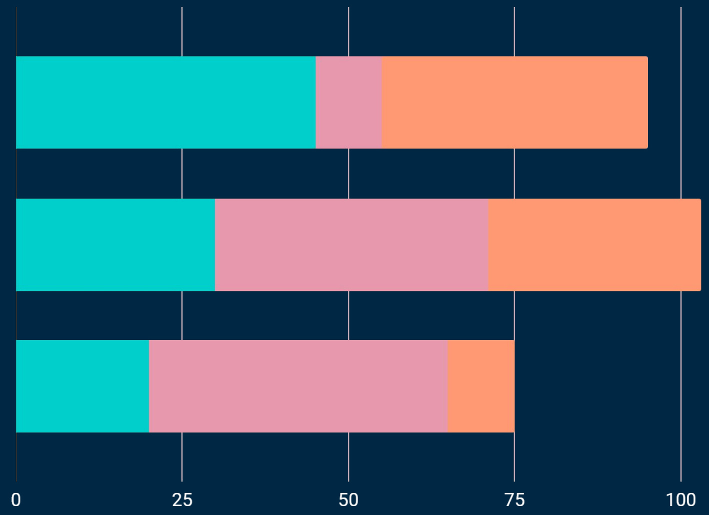
VENUS

Venus is the second planet from the Sun

SATURN

It’s composed mostly of hydrogen and helium

MARKET RESEARCH



NEPTUNE

It's the farthest planet from the Sun

MERCURY

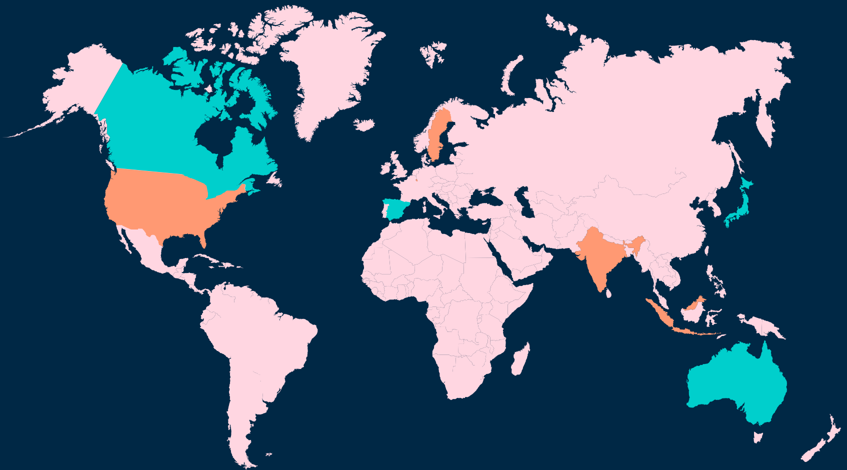
Mercury is the closest planet to the Sun

SATURN

Saturn is composed of hydrogen and helium

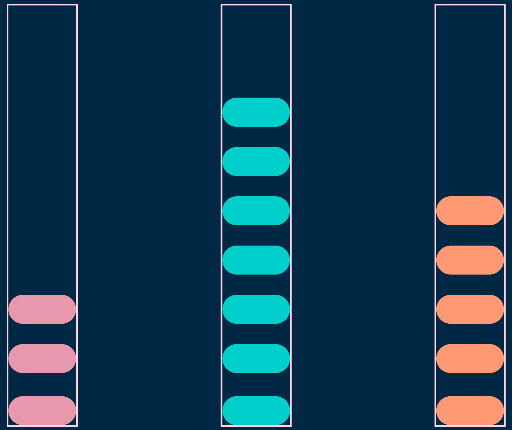
ANALYSIS

OUTREACH



Mars Mercury

TOP RATED VALUES



30% Saturn

80% Neptune

50% Venus

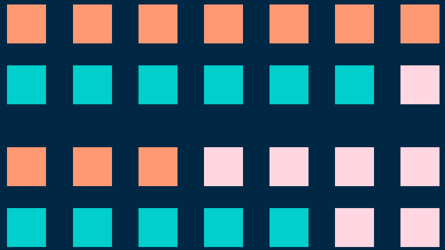
TARGET

GENDER

75% Female

60% Male

AGE



500,000+

Employees

A Picture Is Worth a Thousand Words

"This is a quote. Words full of wisdom that someone important said and can make the reader get inspired."

—SOMEONE FAMOUS

OUR PARTNERS

Venus has an extremely poisonous atmosphere

VENUS

SATURN

Saturn is composed mostly of hydrogen and helium

Despite being red, Mars is actually a cold place

MARS

MERCURY

Mercury is the closest planet to the Sun

TESTIMONIALS

"Mercury is the closest planet to the Sun and the smallest of them all"

—RYAN DIXON

"Venus has a beautiful name and is the second planet from the Sun"

—ALIYA FARLEY

"Jupiter is a gas giant and the biggest planet in the Solar System"

—HENRY McKANE

"Saturn is composed mostly of hydrogen and helium"

—BILLY BROOKS

"The Sun is the star at the center of the Solar System"

—LUCY JADE

"Neptune is the fourth-largest planet in the Solar System"

—ROSE CLARK

AWARDS

	DATE	REASON	DESCRIPTION
MERCURY	2010	Jupiter	It’s the closest planet to the Sun and the smallest one
MARS	2012	Neptune	Despite being red, Mars is actually a cold place
VENUS	2016	Saturn	It has a nice name and is the second planet from the Sun

ALTERNATIVE RESOURCES

RESOURCES

Did you like the resources on this template?
Get them for free at our other websites.

VECTORS

- Abstract pixel rain blue background
- Dashboard element collection
- Red and blue neon fingerprint background
- Dashboard element collection template

PHOTOS

- Female colleagues discussing work at office
- Close-up confident adult woman posing
- Portrait of beautiful adult woman smiling

Instructions for use

In order to use this template, you must credit **Slidesgo** by keeping the **Thanks** slide.

You are allowed to:

- Modify this template.
- Use it for both personal and commercial projects.

You are not allowed to:

- Sublicense, sell or rent any of Slidesgo Content (or a modified version of Slidesgo Content).
- Distribute Slidesgo Content unless it has been expressly authorized by Slidesgo.
- Include Slidesgo Content in an online or offline database or file.
- Offer Slidesgo templates (or modified versions of Slidesgo templates) for download.
- Acquire the copyright of Slidesgo Content.

For more information about editing slides, please read our FAQs or visit Slidesgo School:
https://slidesgo.com/faqs and https://slidesgo.com/slidesgo-school

Fonts & colors used

This presentation has been made using the following fonts:

Share Tech

(https://fonts.google.com/specimen/Share+Tech)

Maven Pro

(https://fonts.google.com/specimen/Maven+Pro)

#002845

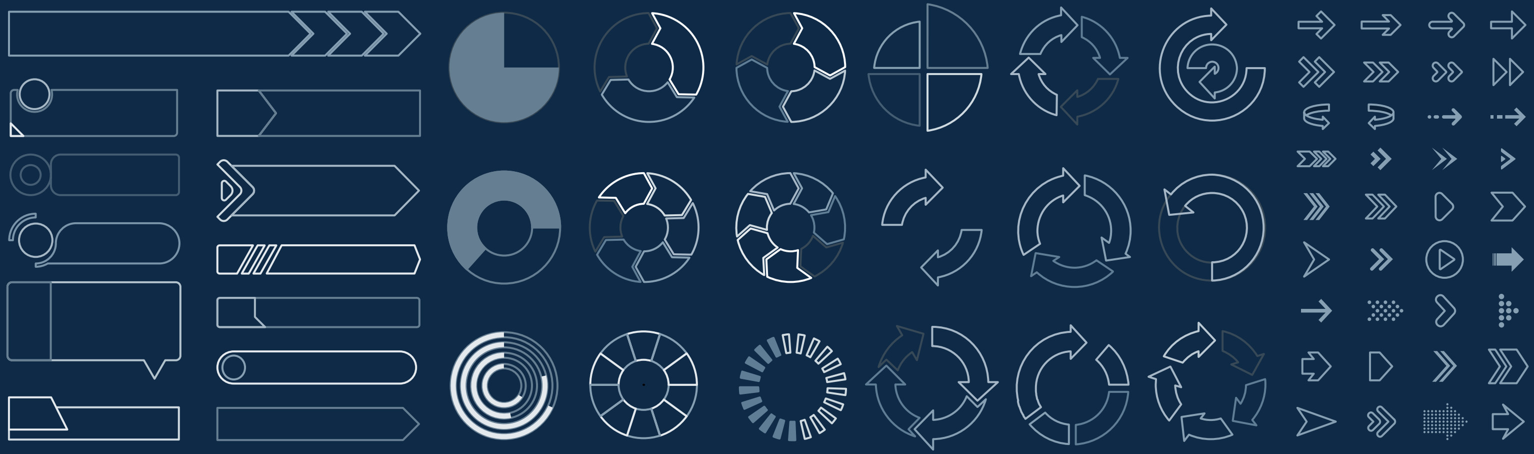
#e898ac

#00cfcc

#ff9973

Use our editable graphic resources...

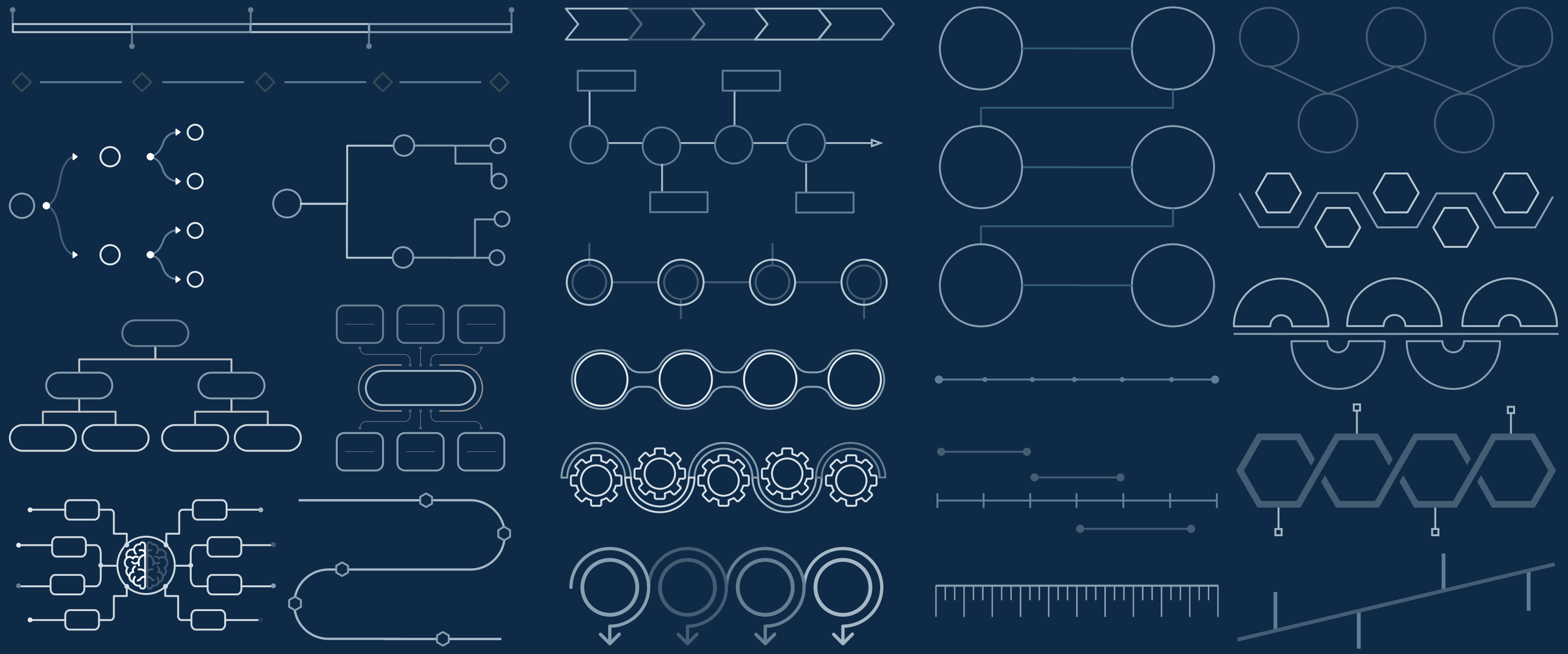
You can easily resize these resources without losing quality. To change the color, just ungroup the resource and click on the object you want to change. Then, click on the paint bucket and select the color you want. Group the resource again when you're done.

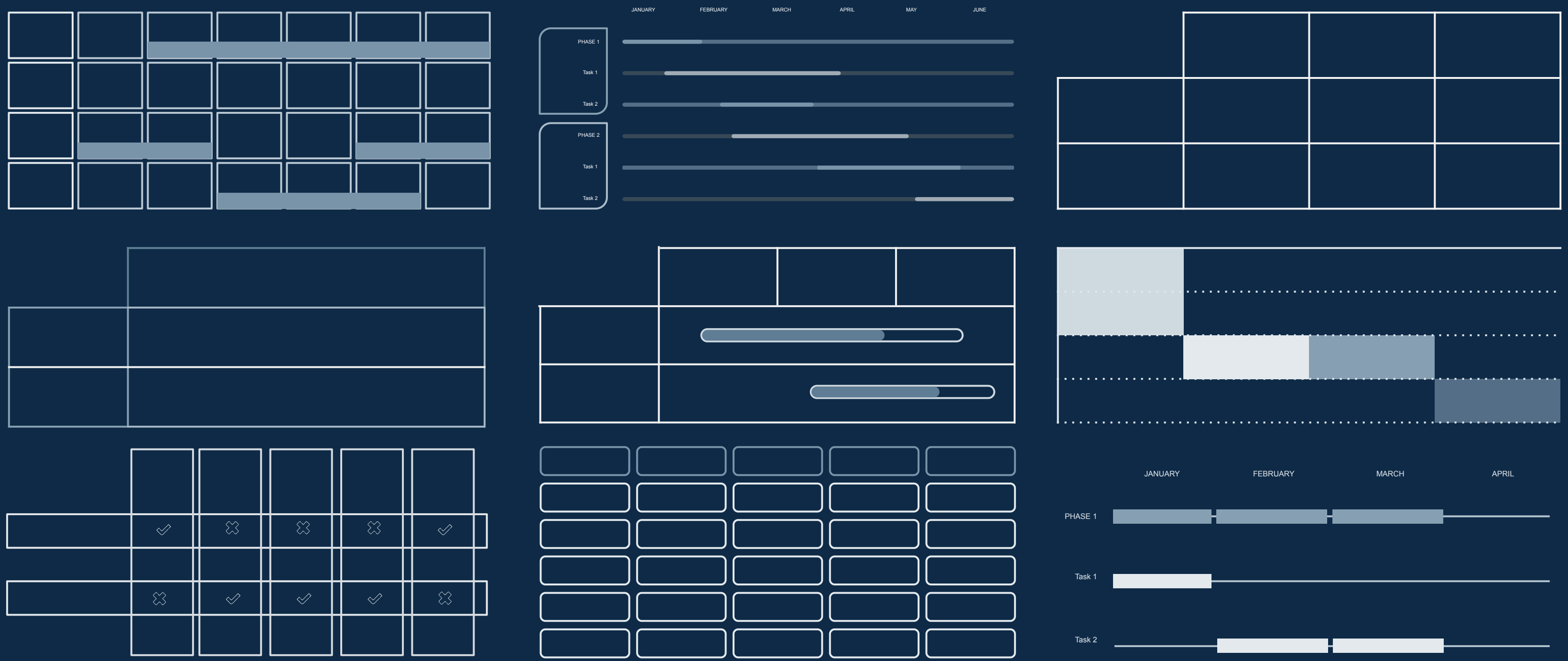


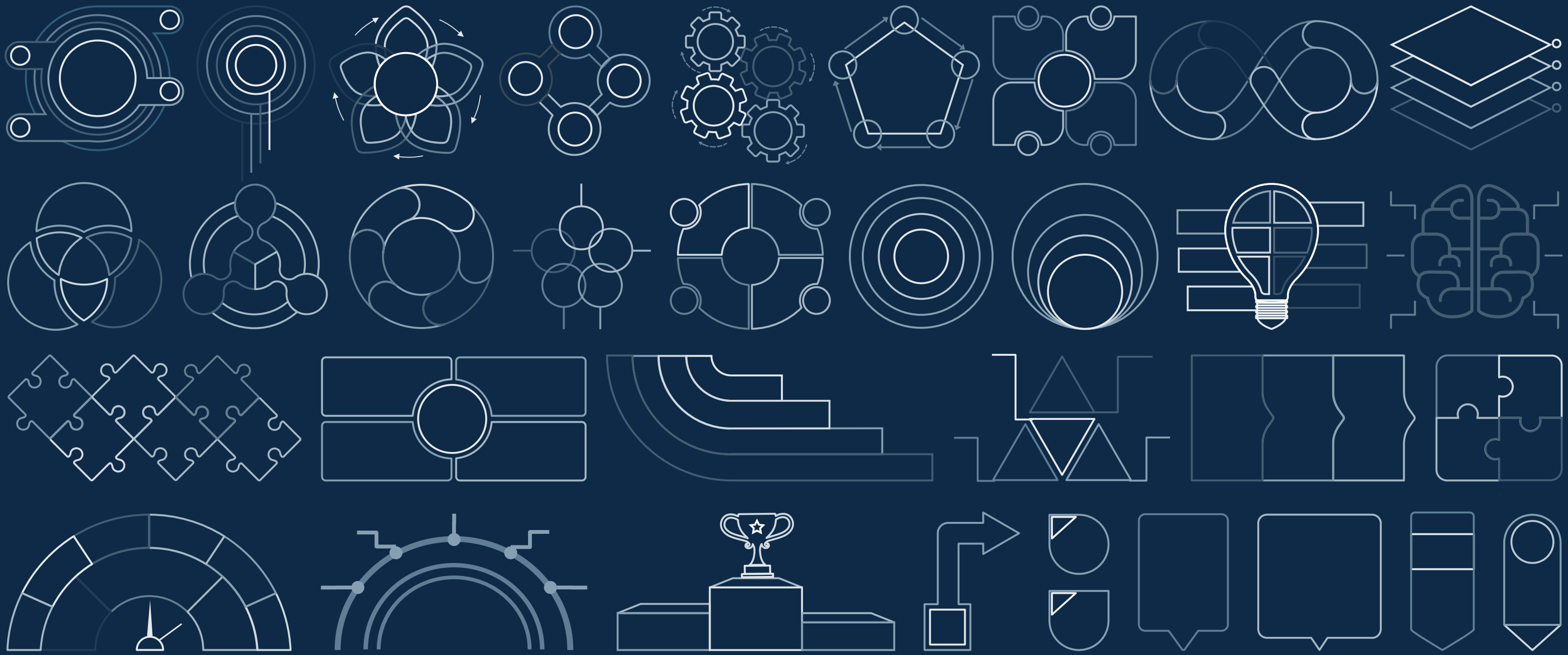


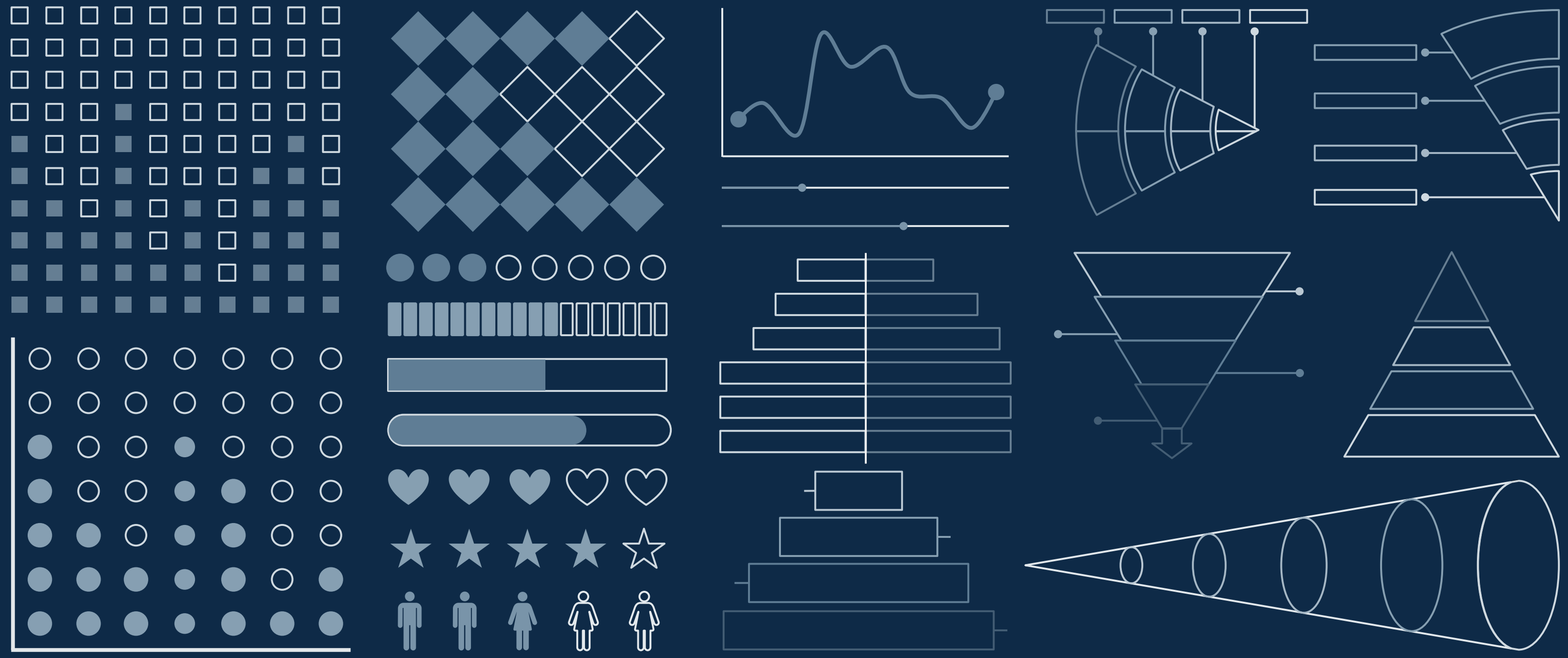







JANUARY
FEBRUARY
MARCH
APRIL
MAY
JUNE
PHASE 1
Task 1
Task 2
PHASE 2
Task 1
Task 2
JANUARY
FEBRUARY
MARCH
APRIL
PHASE 1
Task 1
Task 2





...and our sets of editable icons

You can resize these icons without losing quality.
You can change the stroke and fill color; just select the icon and click on the paint bucket/pen.
In Google Slides, you can also use Flaticon’s extension, allowing you to customize and add even more icons.

Educational Icons

Medical Icons

Business Icons

Teamwork Icons

Help & Support Icons

Avatar Icons

Creative Process Icons

Performing Arts Icons

Nature Icons

SEO & Marketing Icons

slidesgo