













Task Agnostic Dialect Adapters

William Held

Caleb Ziems, Diyi Yang

Dialects of English

- All natural language follows a **systematic** set of rules
- All natural languages experience **variation**
- **dialect**: a group of systematic variations in a language

Country	Total English speakers
 World 	1,179,874,130
 United States 	316,107,532
 India 	128,539,090
 Pakistan 	115,044,691
 Nigeria 	103,198,040

[Rickford 2020](#)

[Various Sources](#)

Example: Negative Inversion

Do these sentences mean the
same thing?

“SHE LOVES THE FACT THAT **NOBODY LIKES** HER.”



“SHE LOVES THE FACT THAT **DON'T NOBODY LIKE** HER.”

[Foreman 1999](#)

Example: Negative Inversion

Do these sentences mean the same thing?

“SHE LOVES THE FACT THAT **NOBODY LIKES** HER.”



“SHE LOVES THE FACT THAT **DON'T NOBODY LIKE** HER.”

[Foreman 1999](#)



High Acceptability



Medium Acceptability



Low Acceptability



Unacceptable

Example: Negative Inversion





Do these sentences mean the same thing?

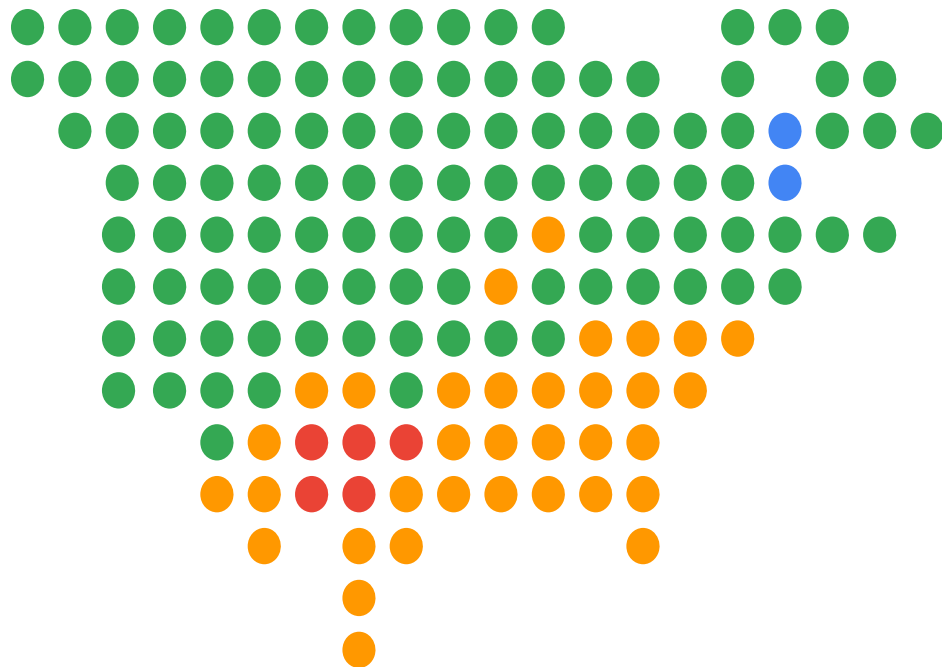
“SHE LOVES THE FACT THAT **NOBODY LIKES** HER.”



“SHE LOVES THE FACT THAT **DON'T NOBODY LIKE** HER.”

[Foreman 1999](#)

-  High Acceptability
-  Medium Acceptability
-  Low Acceptability
-  Unacceptable



[Matyiku, 2011](#)

How does that cause errors?

Hypothesis:

JANE WISHED EVERYONE LOVED HER!

Premises:

“SHE LOVES THAT **NOBODY LIKES HER.”**

“SHE **DOESN'T LOVE THAT **NOBODY LIKES** HER.”**

Learn Boundaries and Embeddings from SAE

Hypothesis:

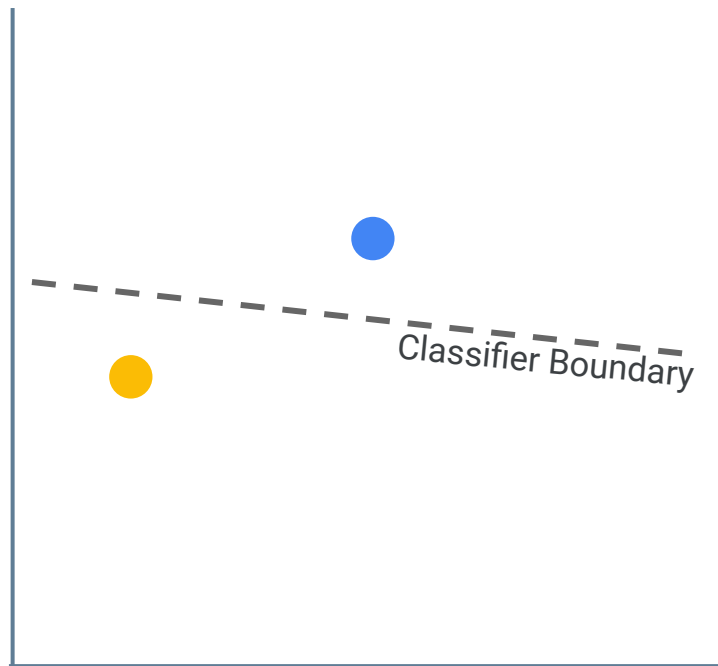
JANE WISHED EVERYONE LOVED HER!

Premises:

“SHE LOVES THAT **NOBODY LIKES HER.**”

“SHE **DOESN'T** LOVE THAT **NOBODY LIKES HER.**”

Sequence Embedding



Same Semantics, Different Embedding

Hypothesis:

JANE WISHED EVERYONE LOVED HER!

Premises:

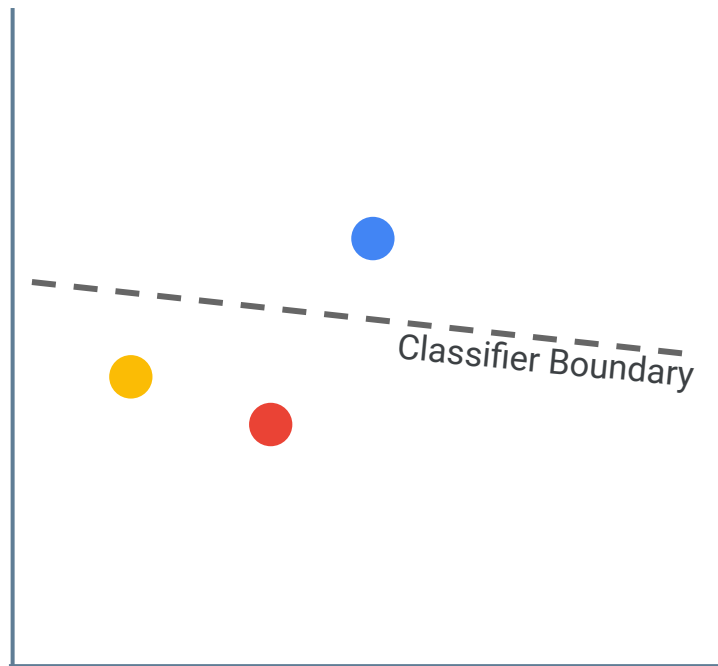
“SHE LOVES THAT **NOBODY LIKES HER.**”



“SHE LOVES THAT **DON'T NOBODY LIKE HER.**”

“SHE **DOESN'T** LOVE THAT **NOBODY LIKES HER.**”

Sequence Embedding



Retraining Gives New Boundaries and New Embeddings!

Hypothesis:
JANE WISHED EVERYONE LOVED HER!

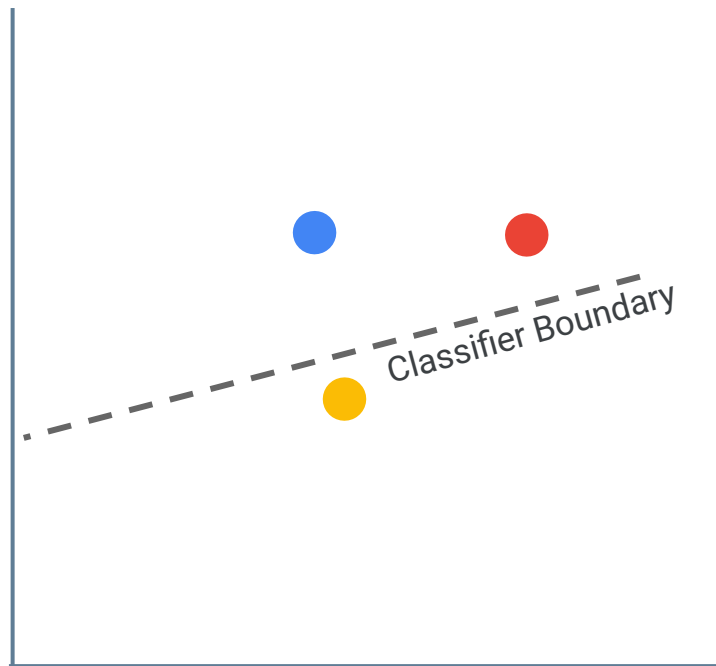
Premises:
"SHE LOVES THAT **NOBODY LIKES** HER."



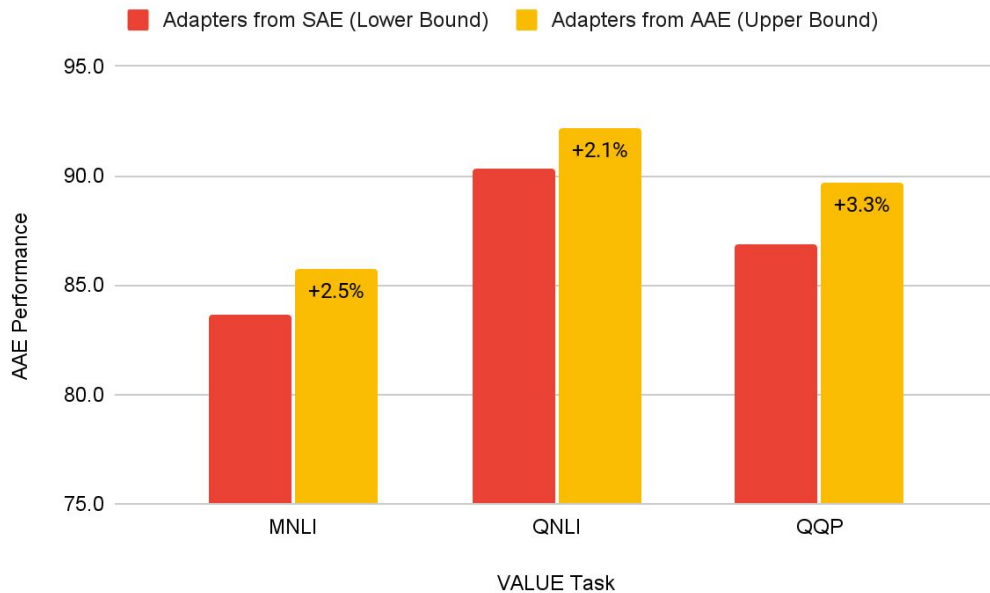
"SHE LOVES THAT **DON'T NOBODY LIKE** HER."

"SHE **DOESN'T** LOVE THAT **NOBODY LIKES** HER."

Sequence Embedding



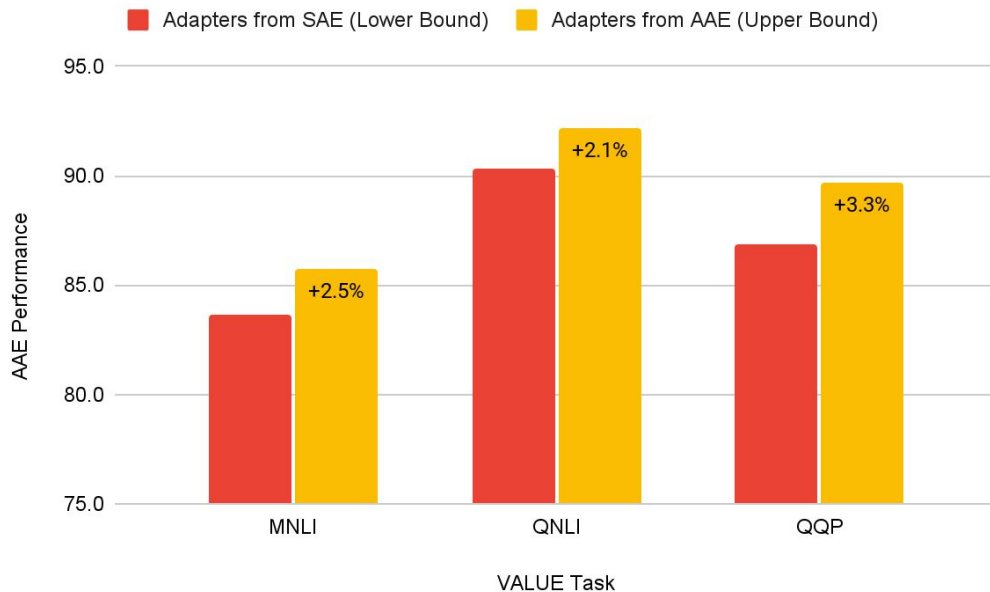
Retraining Works!



Reproduce VALUE Experiments!



Cost of Retraining is Linear With Number of Tasks

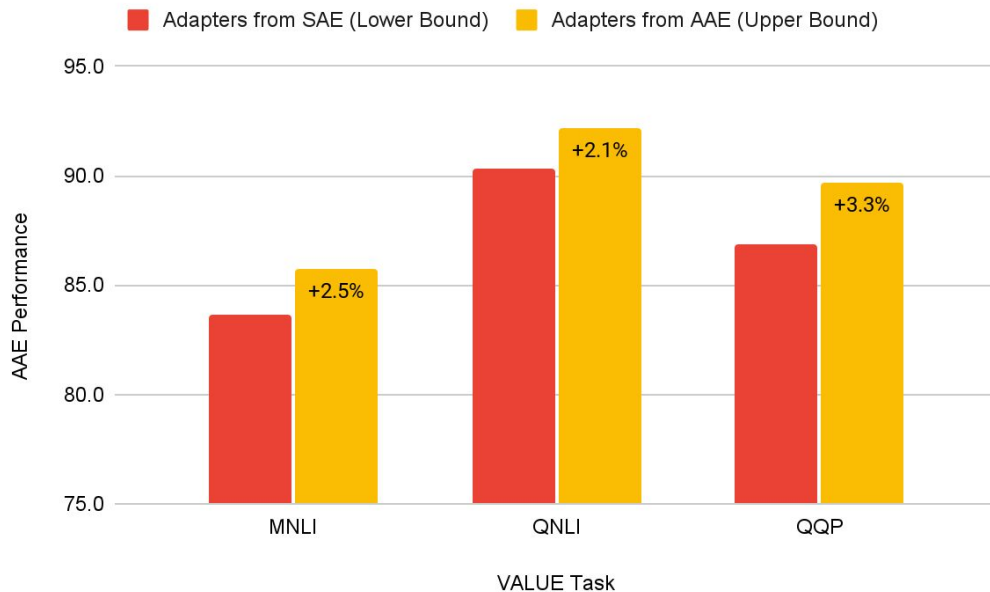


T * Cost

Reproduce VALUE Experiments!



Cost of Retraining is Linear With Number of Tasks



T * Cost **Carbon Emissions**
\$\$\$ on GPU Hours
Parameters Stored

Reproduce VALUE Experiments!



Same Boundaries, New Embeddings

Hypothesis:

JANE WISHED EVERYONE LOVED HER!

Premises:

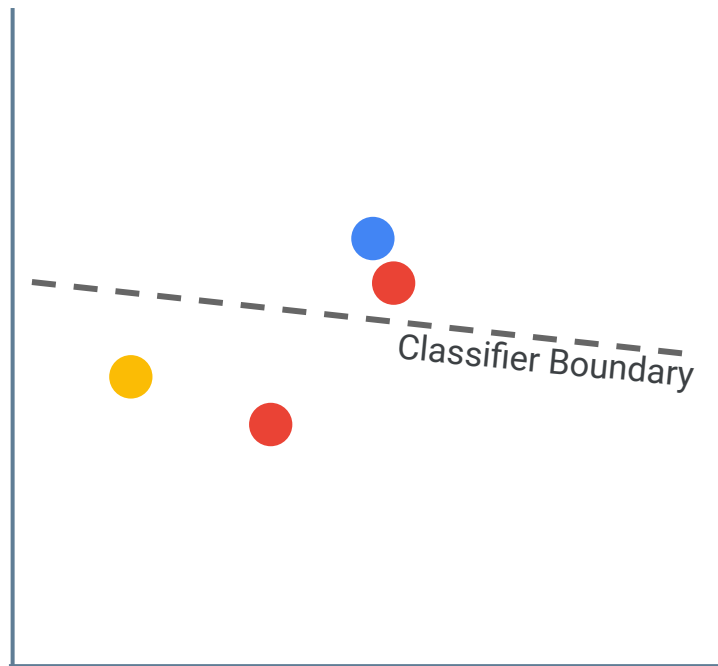
“SHE LOVES THAT **NOBODY LIKES HER.**”



“SHE LOVES THAT **DON'T NOBODY LIKE HER.**”

“SHE **DOESN'T** LOVE THAT **NOBODY LIKES HER.**”

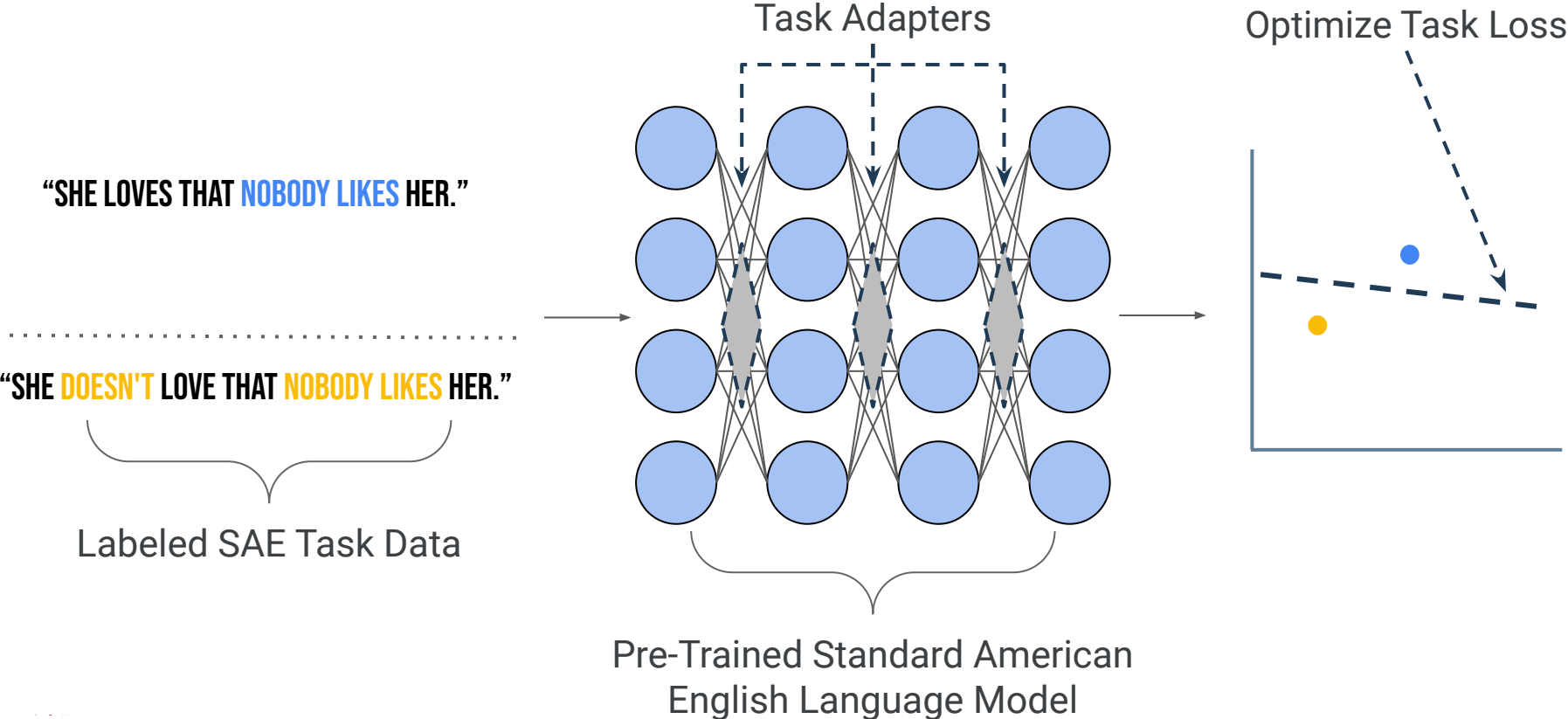
Sequence Embedding



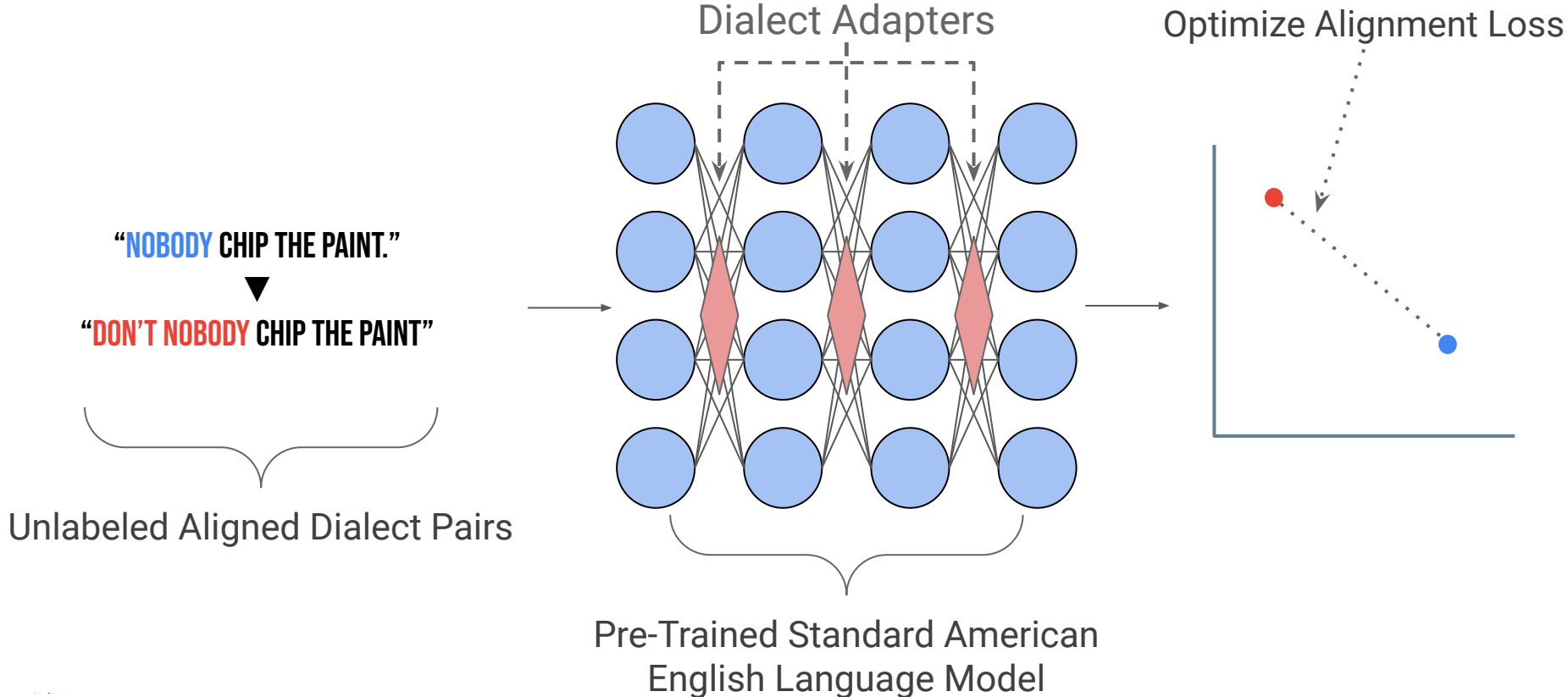
TADA: Task Agnostic Dialect Adaptation for English

- **Goals:** Provide Plug-and-Play Dialect Robustness For All Tasks at Constant Cost
- **Focus:** Align Dialect Embeddings to Standard American English Embeddings
- **Methods:**
 1. Contrastive Alignment at the Sequence Level
 2. Adversarial Alignment at the Token Level
- **Evaluation:**
 1. Multi-VALUE Transformed GLUE
 - 7 Common NLP Classification Tasks
 - 4 Non-SAE Dialects
 - African-American, Indian, Nigerian, and Singaporean Englishes

Traditional Task Adapter Training



Task Agnostic Dialect Adapter Training



What is “Alignment Loss”?



Alignment Loss

$$\| \text{CLS}_{\text{SAE}} - \text{CLS}_{\text{Dial}} \|_2$$

Minimize Euclidean Distance between
Classification token of Target Dialect
and Standard American English

Alignment Type

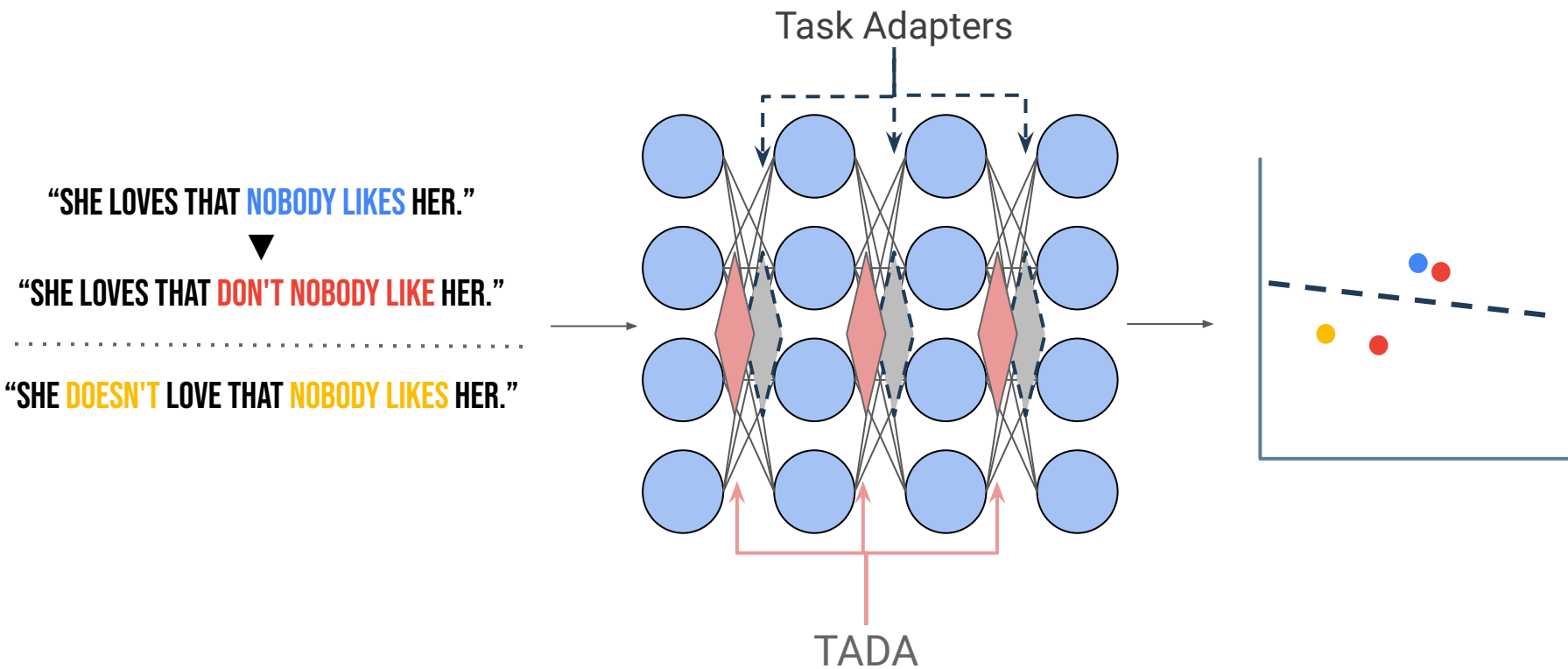
Sequence

$$- \text{Critic}(\text{Dial})$$

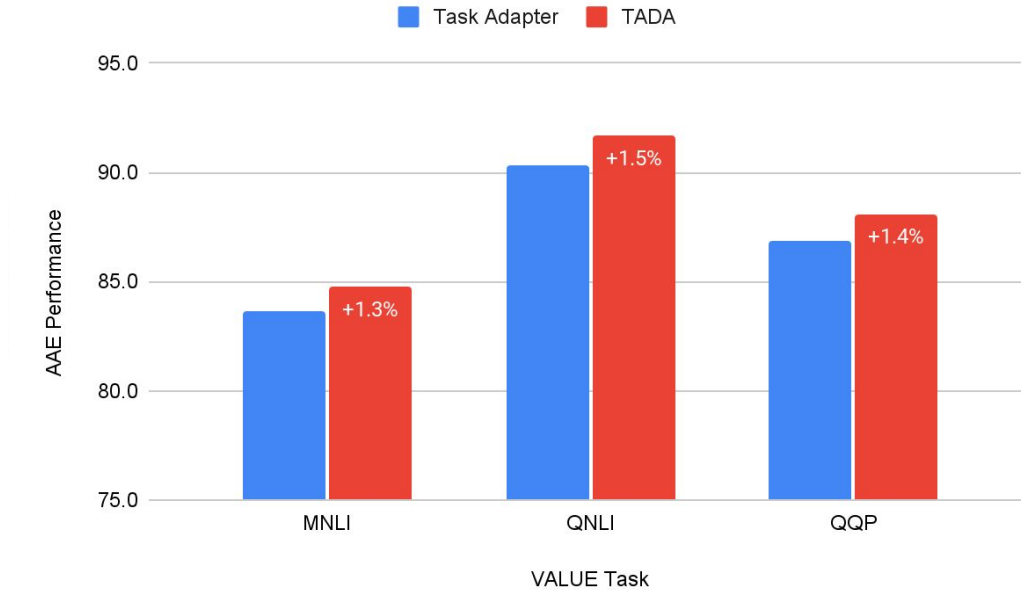
Maximize the Score Given to a
Dialectal Sentence By an Adversarial
Transformer

Token

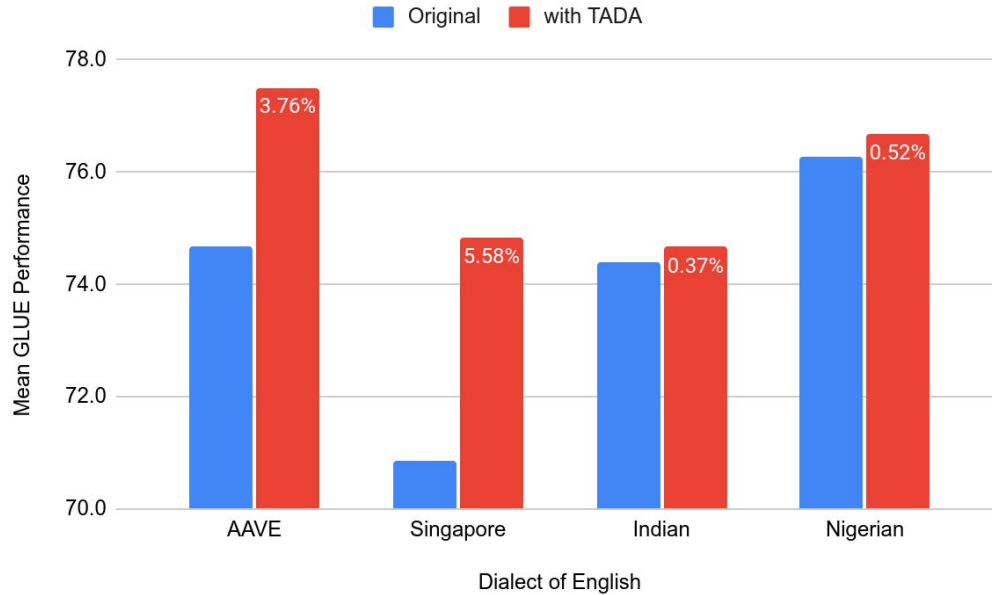
Start with Task Adapters, Add Tada, and...



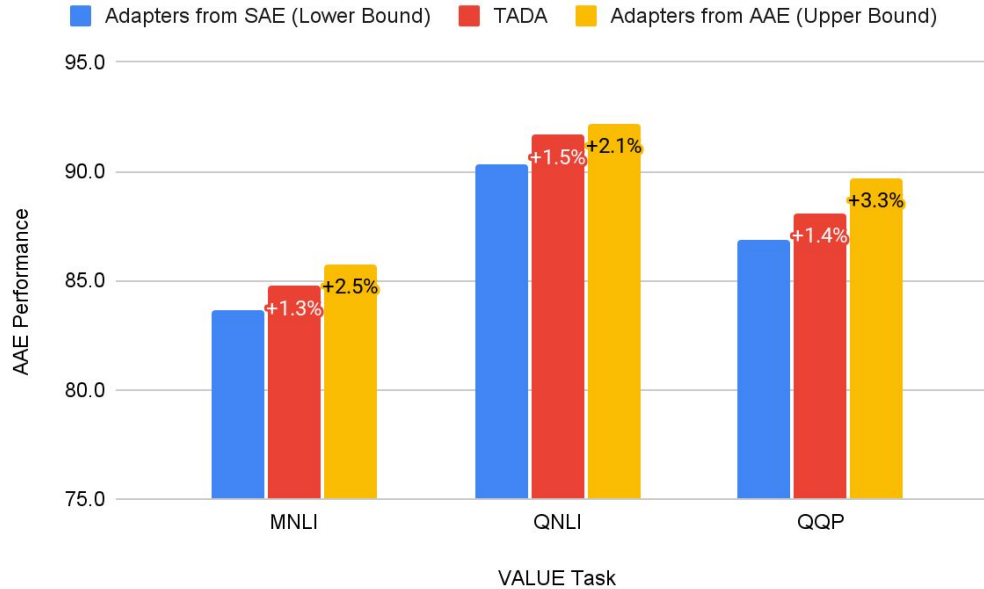
Improve Dialect Performance at Constant Cost!



Works for Multiple Dialects!



There is room-for-growth vs. Task-Specific Methods



Foundations for Task-Agnostic Dialect Research

Dialect Contrastive
Learning Code



Dialect Adapters on 🙌



TADA Paper

Upon Request!
(Anonymity Prevents Arxiv)

I want your input!

Comments on this doc are open for discussion!

OR

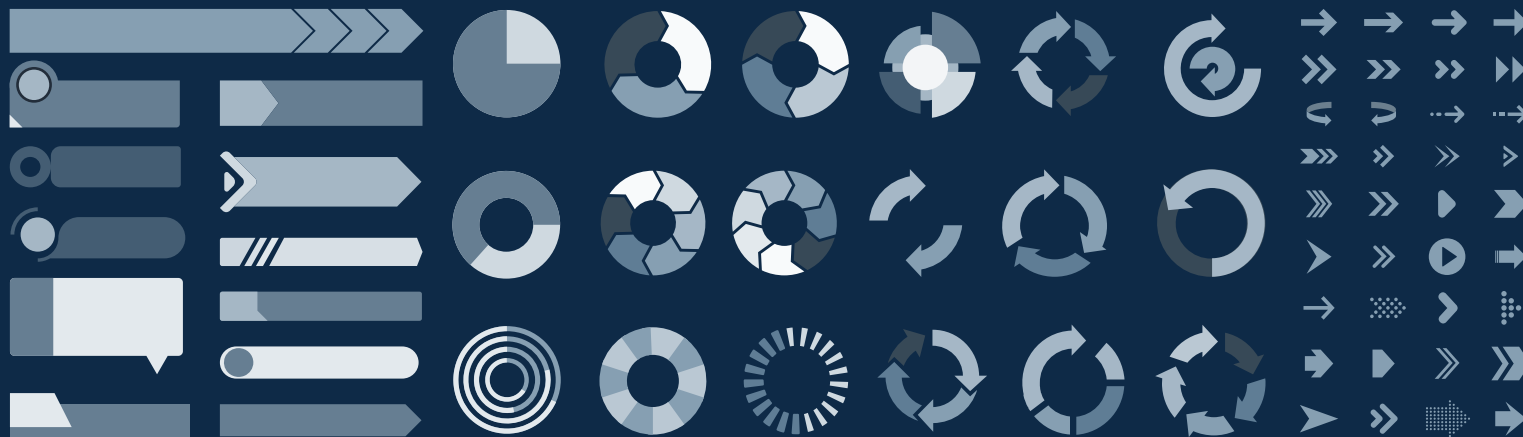
Fill out my anonymous feedback form

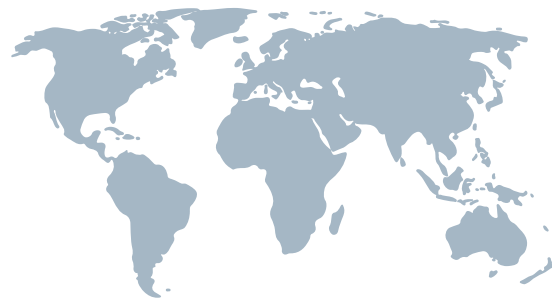


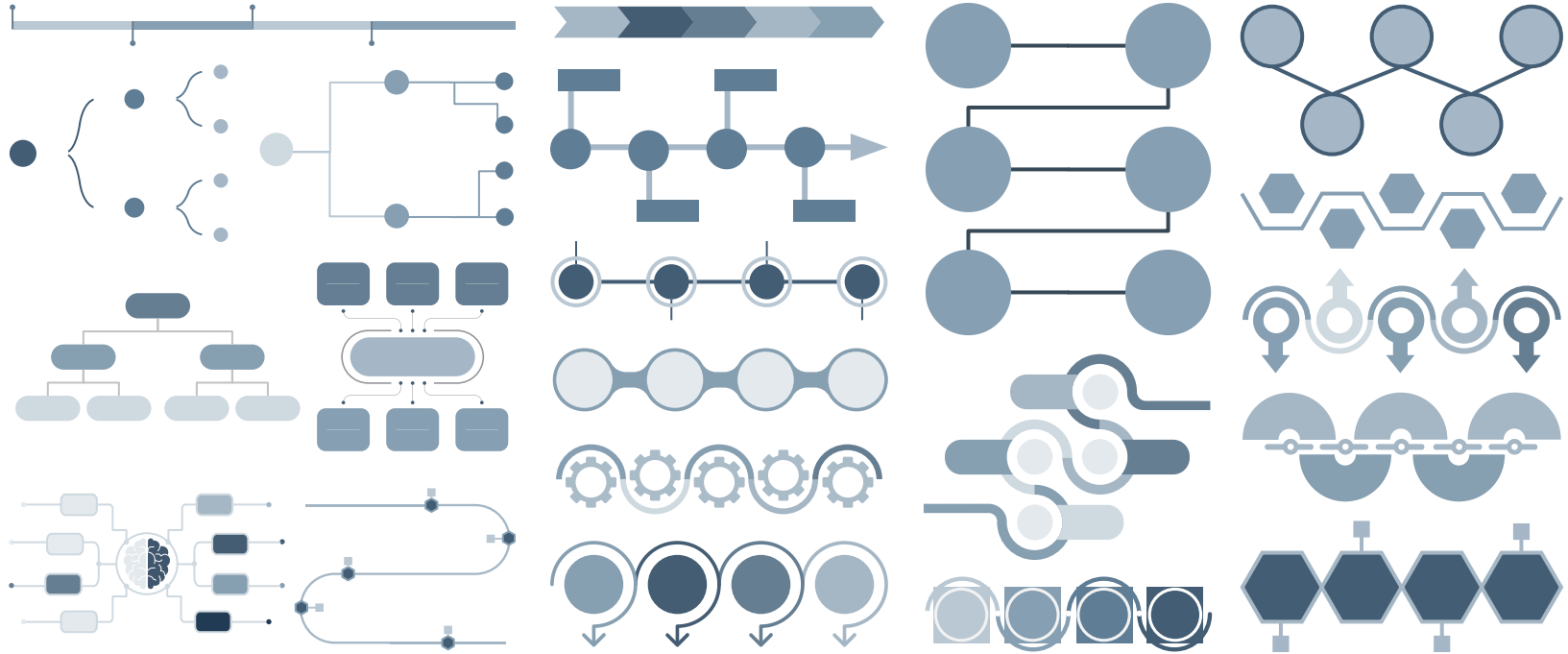
Use our editable graphic resources...

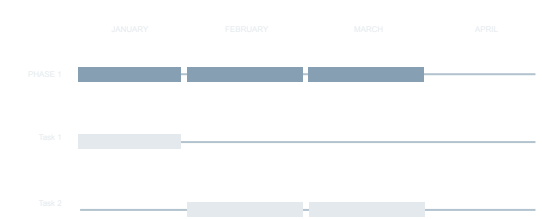
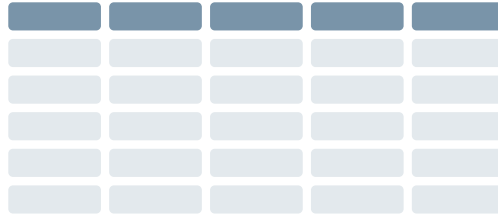
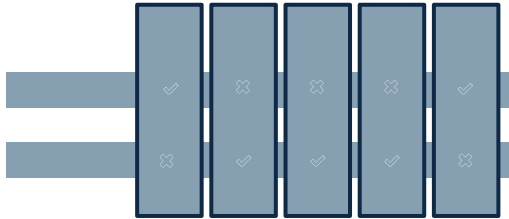
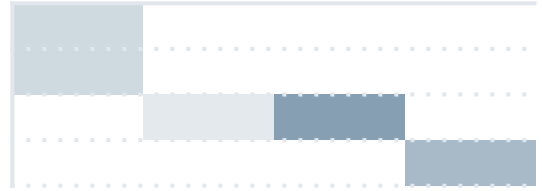
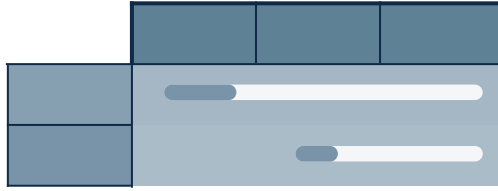
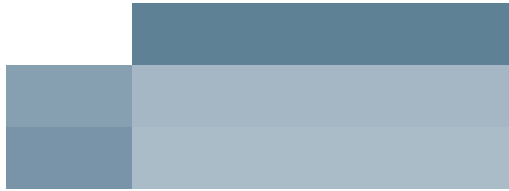
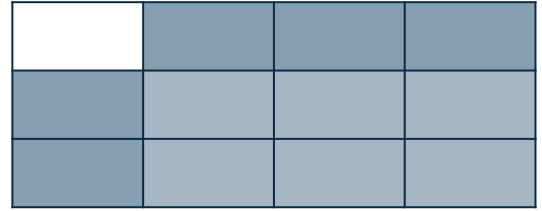
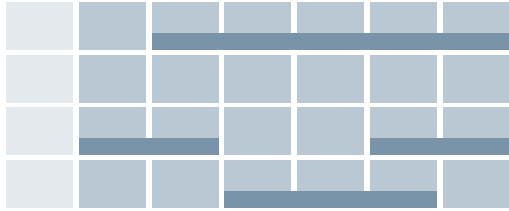
You can easily *resize* these resources without losing quality. To *change the color*, just ungroup the resource and click on the object you want to change. Then, click on the paint bucket and select the color you want.

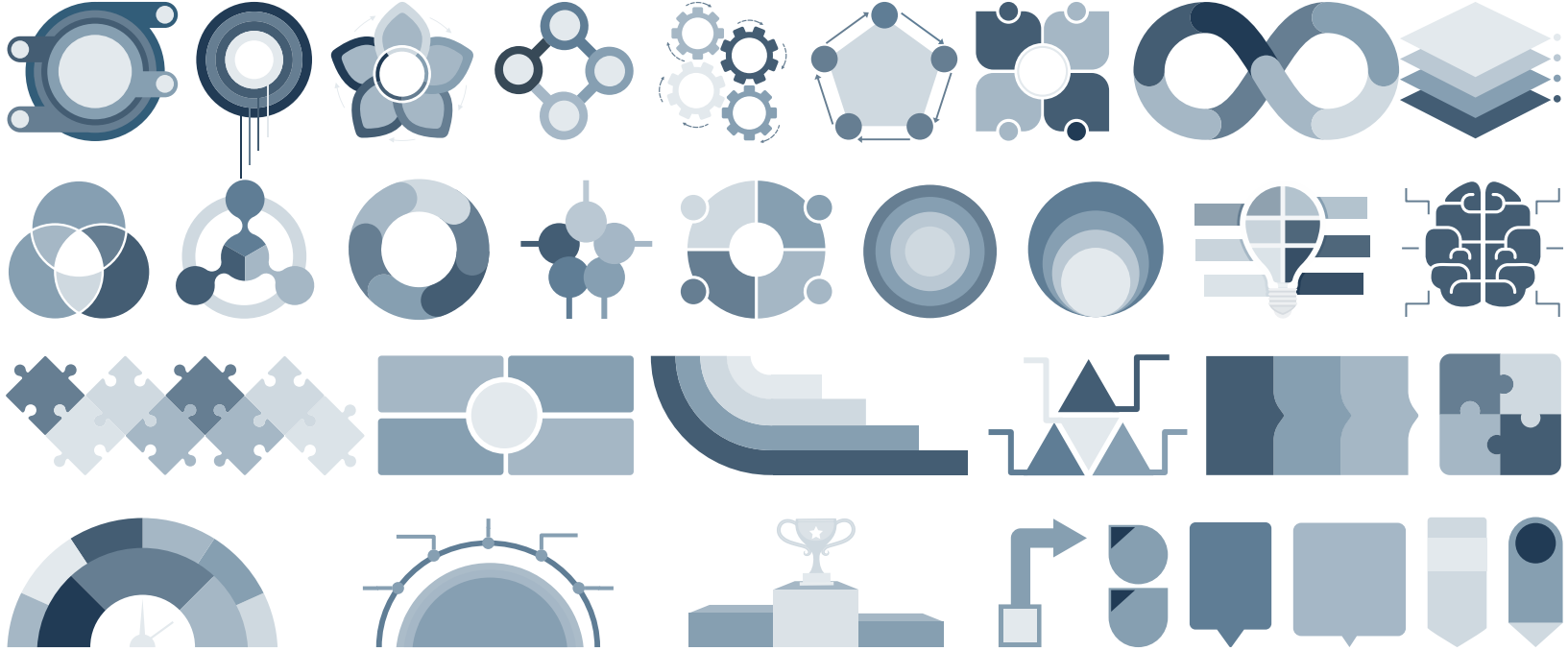
Group the resource again when you're done. You can also look for more *infographics* on [Slidesgo](#).

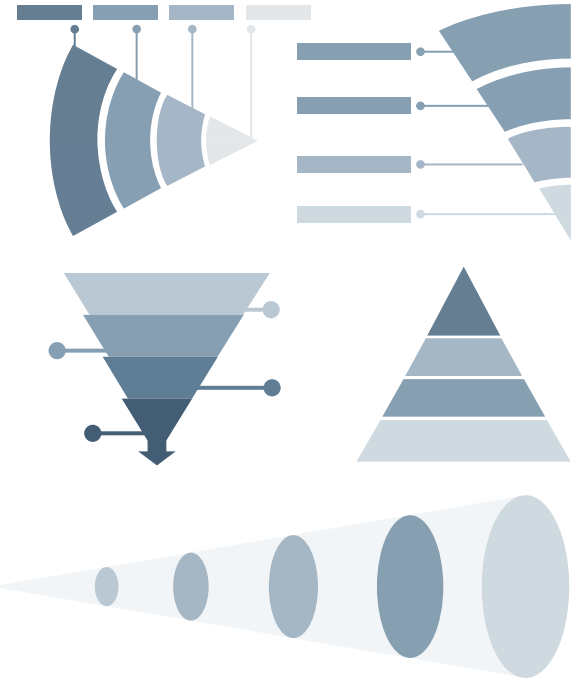
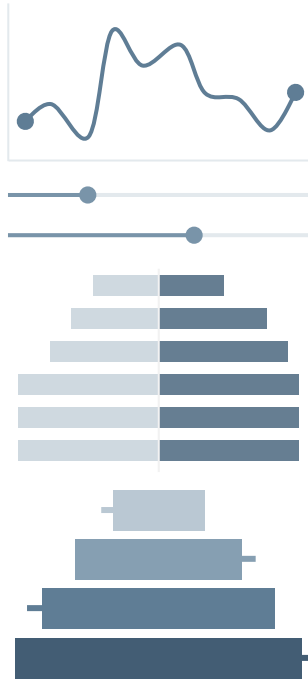
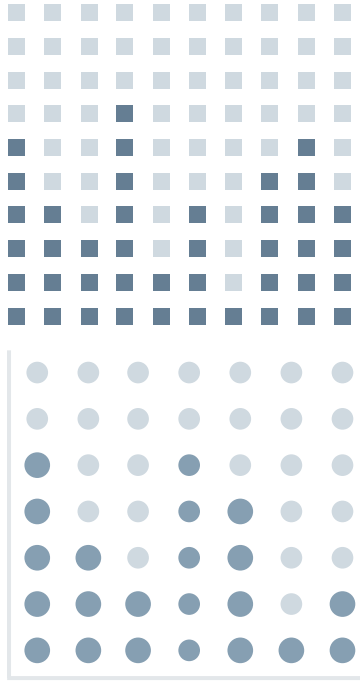












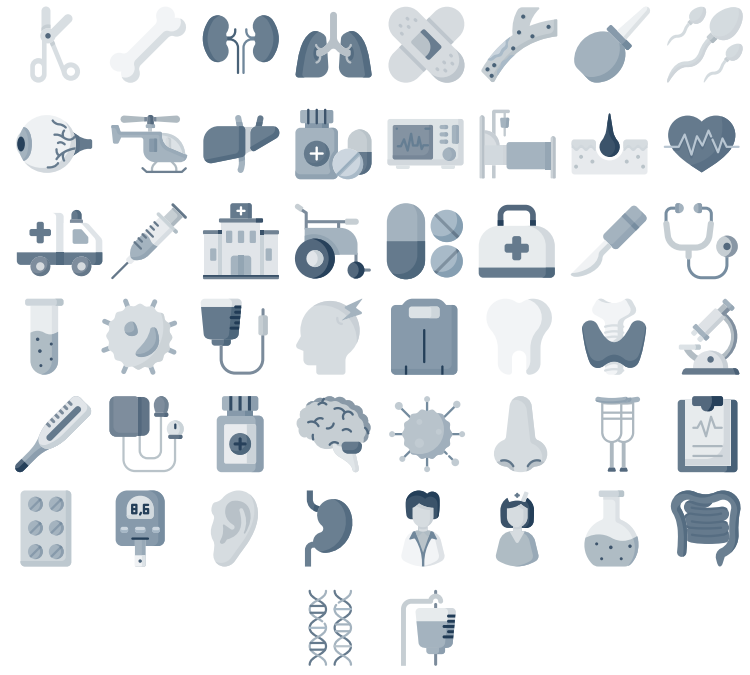
...and our sets of editable icons

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.





Help & Support Icons



Avatar Icons



Creative Process Icons



Performing Arts Icons



Nature Icons



