



# Maximizing Multimodal

**Nathan "Nate" Day**  
*& the Data Wranglers*

2025-04-23  
Charlottesville, VA, USA





# The Data Wranglers



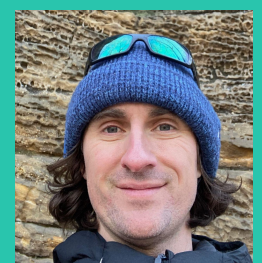
Adam Gamble - 🧑🏻 👑



Kennon Ballou - 🧑🏻🔧 🛠️



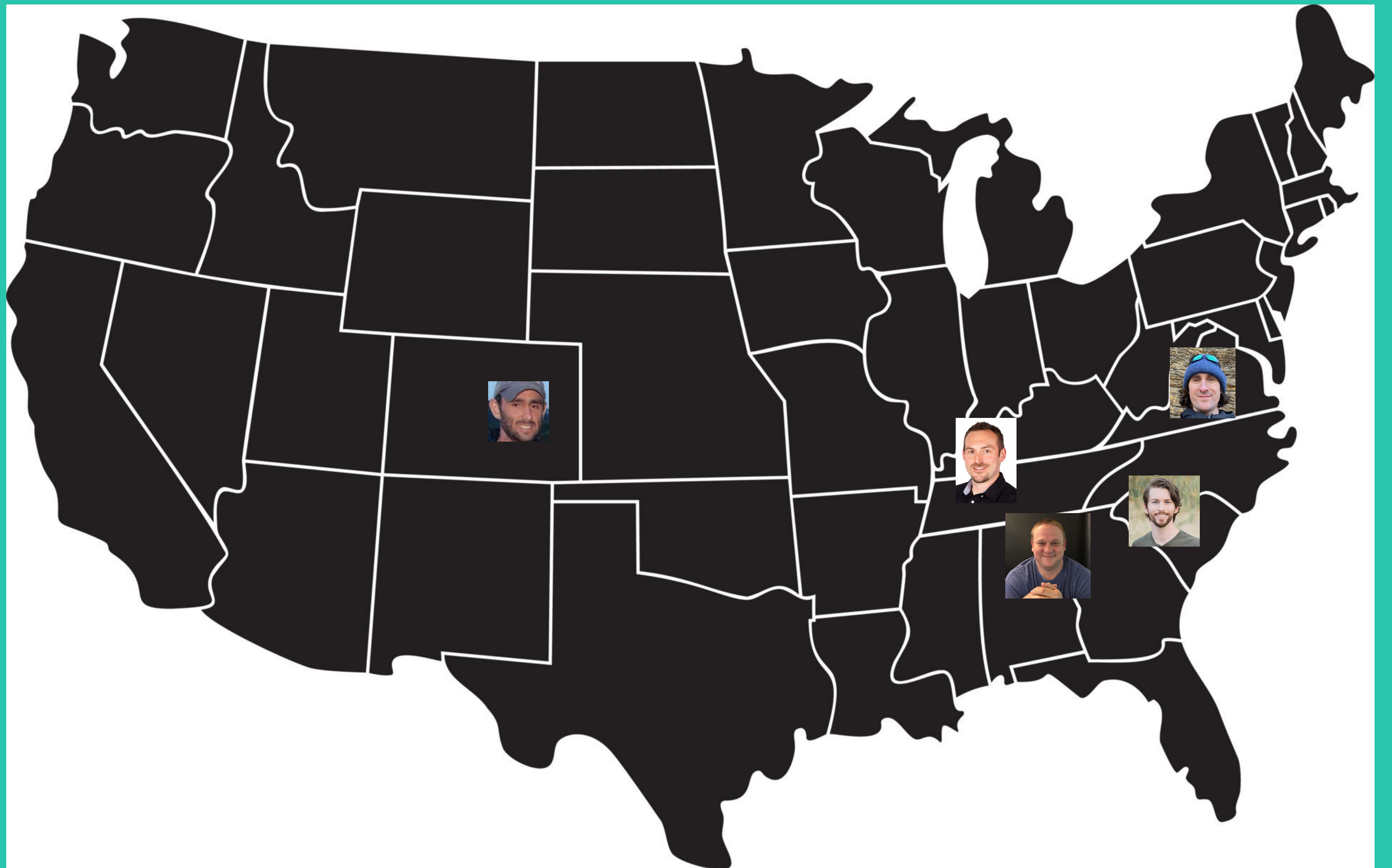
Ryan Gerres - 📊 🧙🏻



Me - 🔍 🧑🏻🔬



Chad Marshall - 🕵🏻 💰

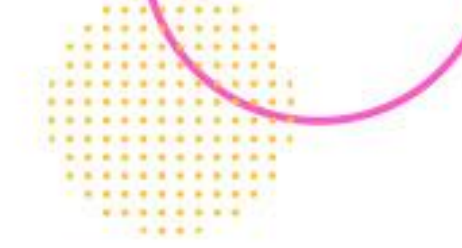






Why we're here

**To inspire and  
simplify the  
creative  
process**





Customers & users

# Who are our customers



## Freelancers

Advanced designers using us for as-needed projects. Will pay for images when needed.



## Career Designers

Professional designers who are using us on a daily basis for work. Willing to subscribe to Pro, possibly as part of a team.



## Hobbyists & Crafters

Beginners who need free graphics and images on a weekly basis for crafting projects and hobby projects.



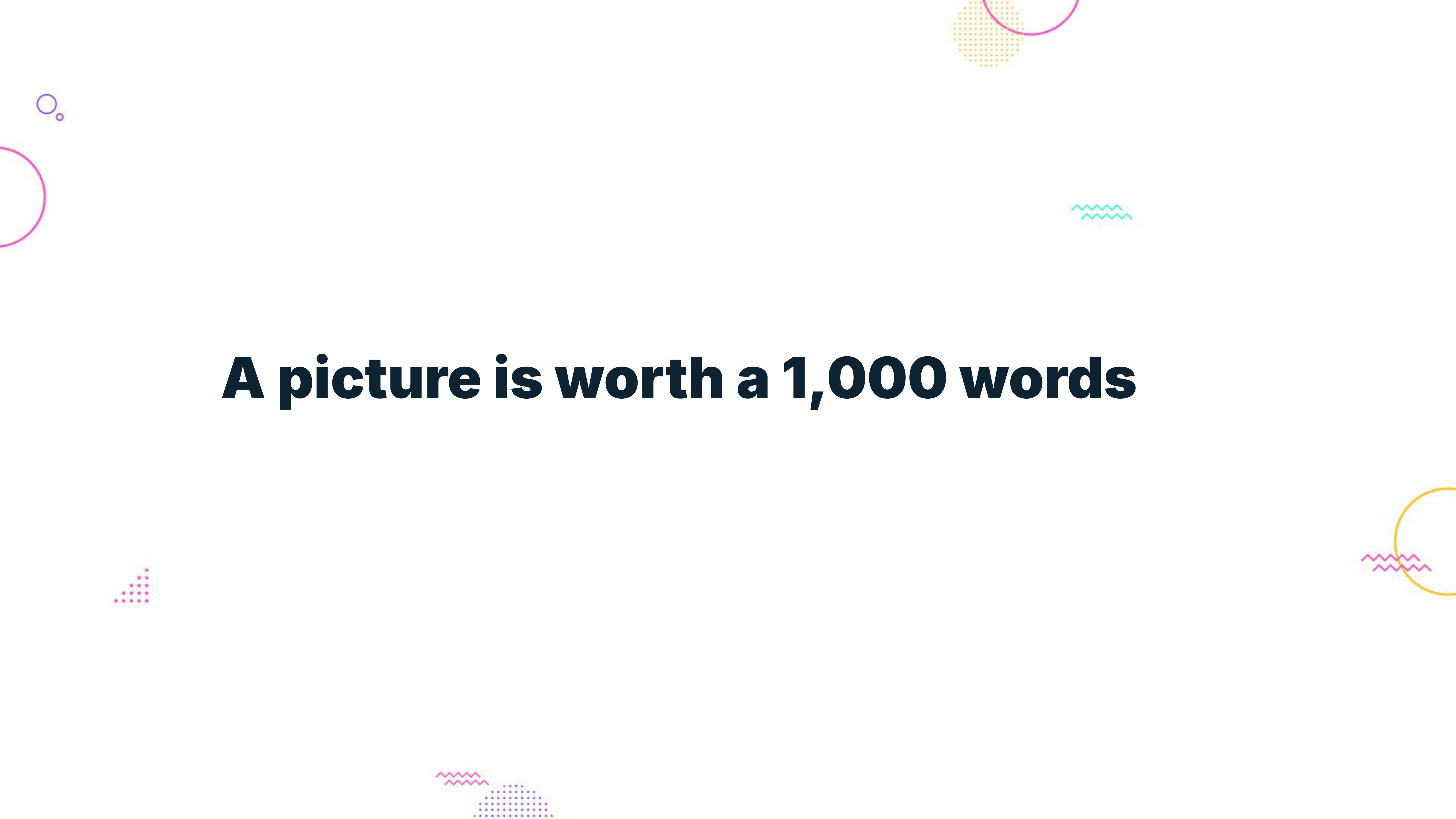
## Entrepreneurs & Brands

Intermediate design skills using our service for monthly deliverables, including print and digital projects.

Our plan

# Champion creativity in every person





**A picture is worth a 1,000 words**



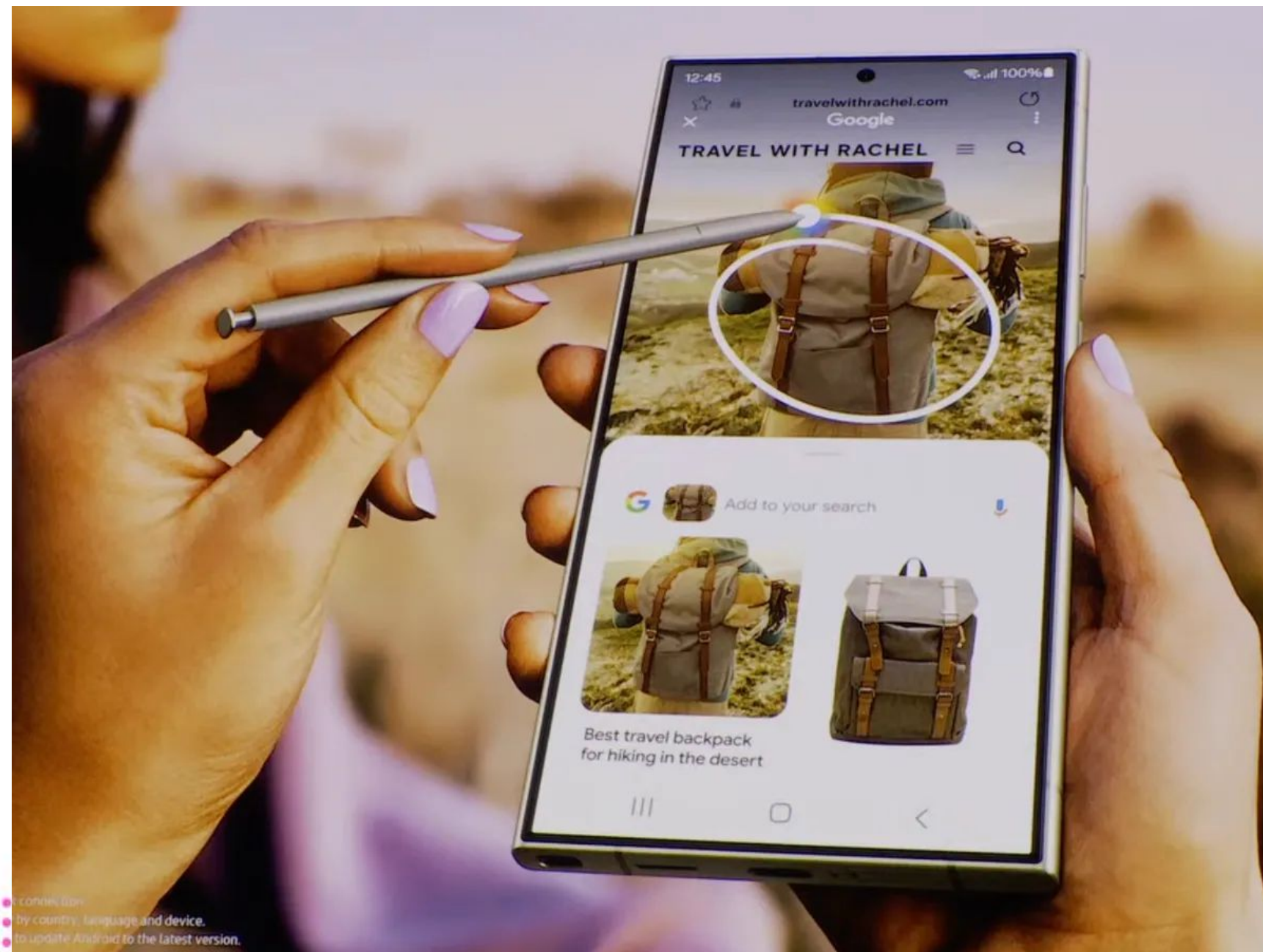


# **A picture is worth a 1,000 words**

So why are we all still searching with text?



# Circle to Search





# Talk Timeline

February 2023

April 2025



Circle-to-Search  
January 2024

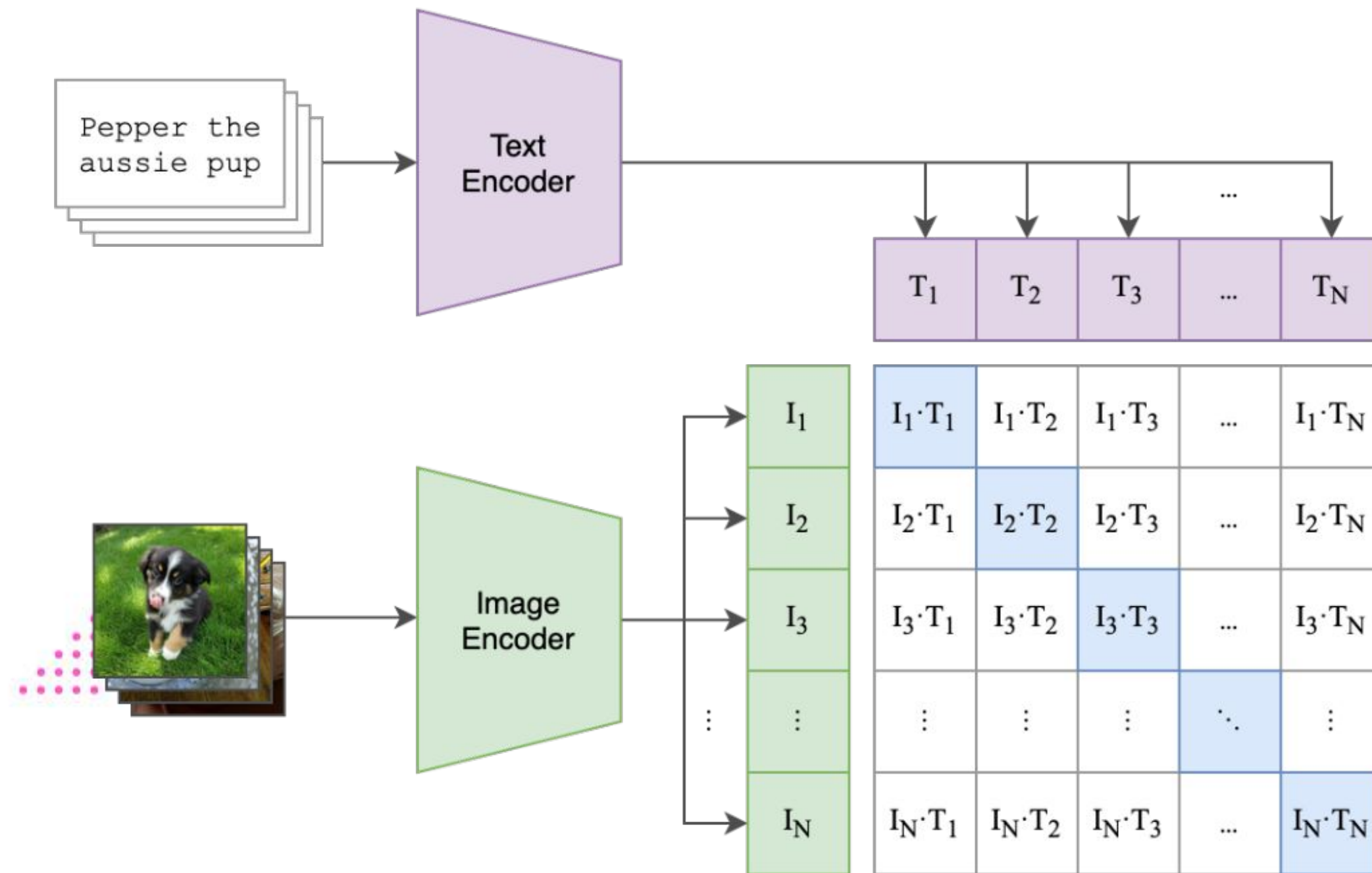


You are Here ^

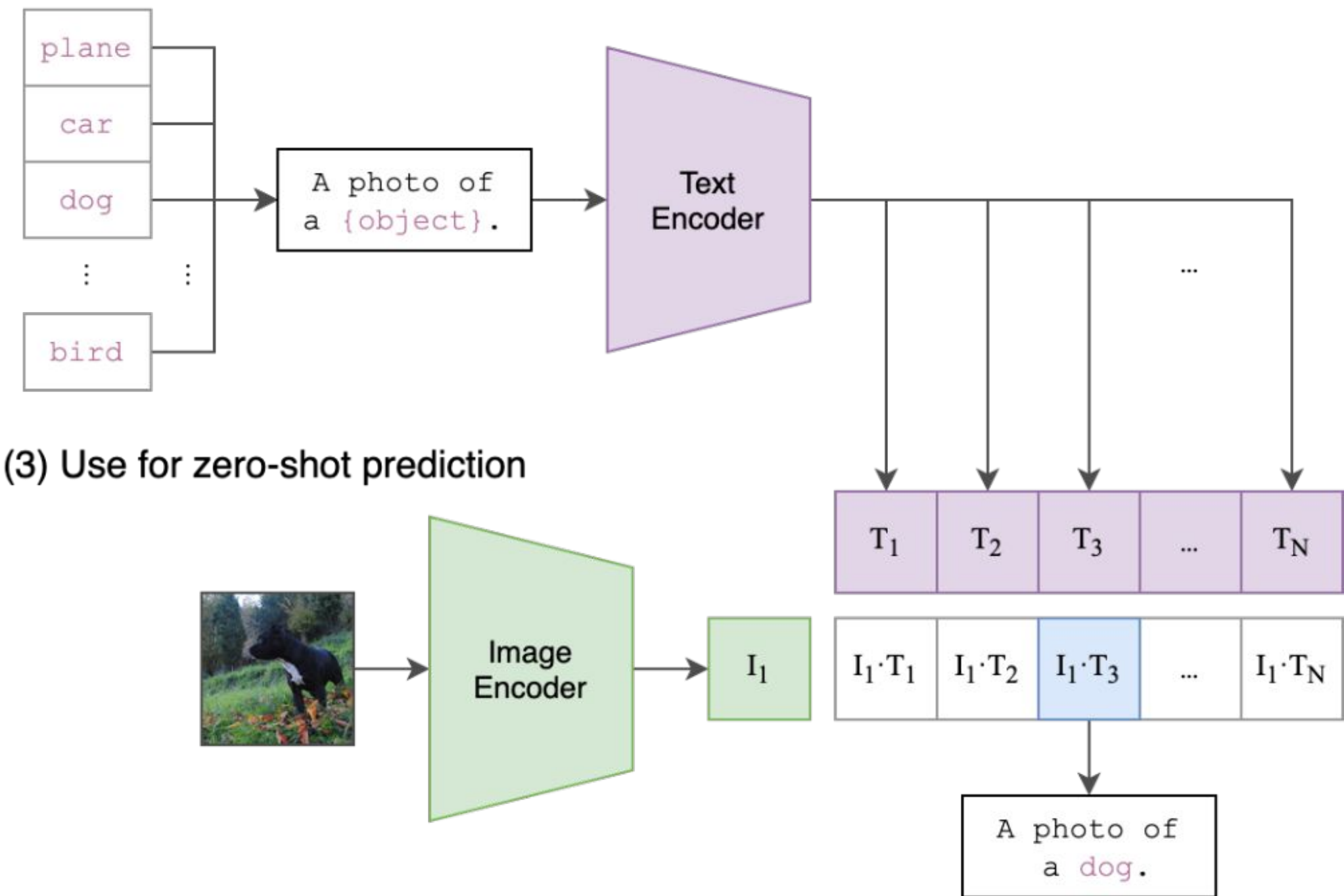


# CLIP

## (1) Contrastive pre-training



## (2) Create dataset classifier from label text



## (3) Use for zero-shot prediction

# CLIP

A powerful zero-shot classifier

|                | aYahoo      | ImageNet    | SUN         |
|----------------|-------------|-------------|-------------|
| Visual N-Grams | 72.4        | 11.5        | 23.0        |
| CLIP           | <b>98.4</b> | <b>76.2</b> | <b>58.5</b> |



# openCLIP

| Model                                                | Training data | Resolution | # of samples seen | ImageNet zero-shot acc. |
|------------------------------------------------------|---------------|------------|-------------------|-------------------------|
| ConvNext-Base                                        | LAION-2B      | 256px      | 13B               | 71.5%                   |
| ConvNext-Large                                       | LAION-2B      | 320px      | 29B               | 76.9%                   |
| ConvNext-XXLarge                                     | LAION-2B      | 256px      | 34B               | 79.5%                   |
| ViT-B/32                                             | DataComp-1B   | 256px      | 34B               | 72.8%                   |
| ViT-B/16                                             | DataComp-1B   | 224px      | 13B               | 73.5%                   |
| ViT-L/14                                             | LAION-2B      | 224px      | 32B               | 75.3%                   |
| ViT-H/14                                             | LAION-2B      | 224px      | 32B               | 78.0%                   |
| ViT-L/14                                             | DataComp-1B   | 224px      | 13B               | 79.2%                   |
| ViT-G/14                                             | LAION-2B      | 224px      | 34B               | 80.1%                   |
|                                                      |               |            |                   |                         |
| ViT-L/14-quickgelu ( <a href="#">Original CLIP</a> ) | WIT           | 224px      | 13B               | 75.5%                   |
| ViT-SO400M/14 ( <a href="#">SigLIP</a> )             | WebLI         | 224px      | 45B               | 82.0%                   |
| ViT-L/14 ( <a href="#">DFN</a> )                     | DFN-2B        | 224px      | 39B               | 82.2%                   |
| ViT-SO400M-14-SigLIP-384 ( <a href="#">SigLIP</a> )  | WebLI         | 384px      | 45B               | 83.1%                   |
| ViT-H/14-quickgelu ( <a href="#">DFN</a> )           | DFN-5B        | 224px      | 39B               | 83.4%                   |
| ViT-H-14-378-quickgelu ( <a href="#">DFN</a> )       | DFN-5B        | 378px      | 44B               | 84.4%                   |

# openCLIP

**OpenAI's CLIP**  
January 2021

**Google's SigCLIP**  
September 2023

**openCLIP**  
December 2022

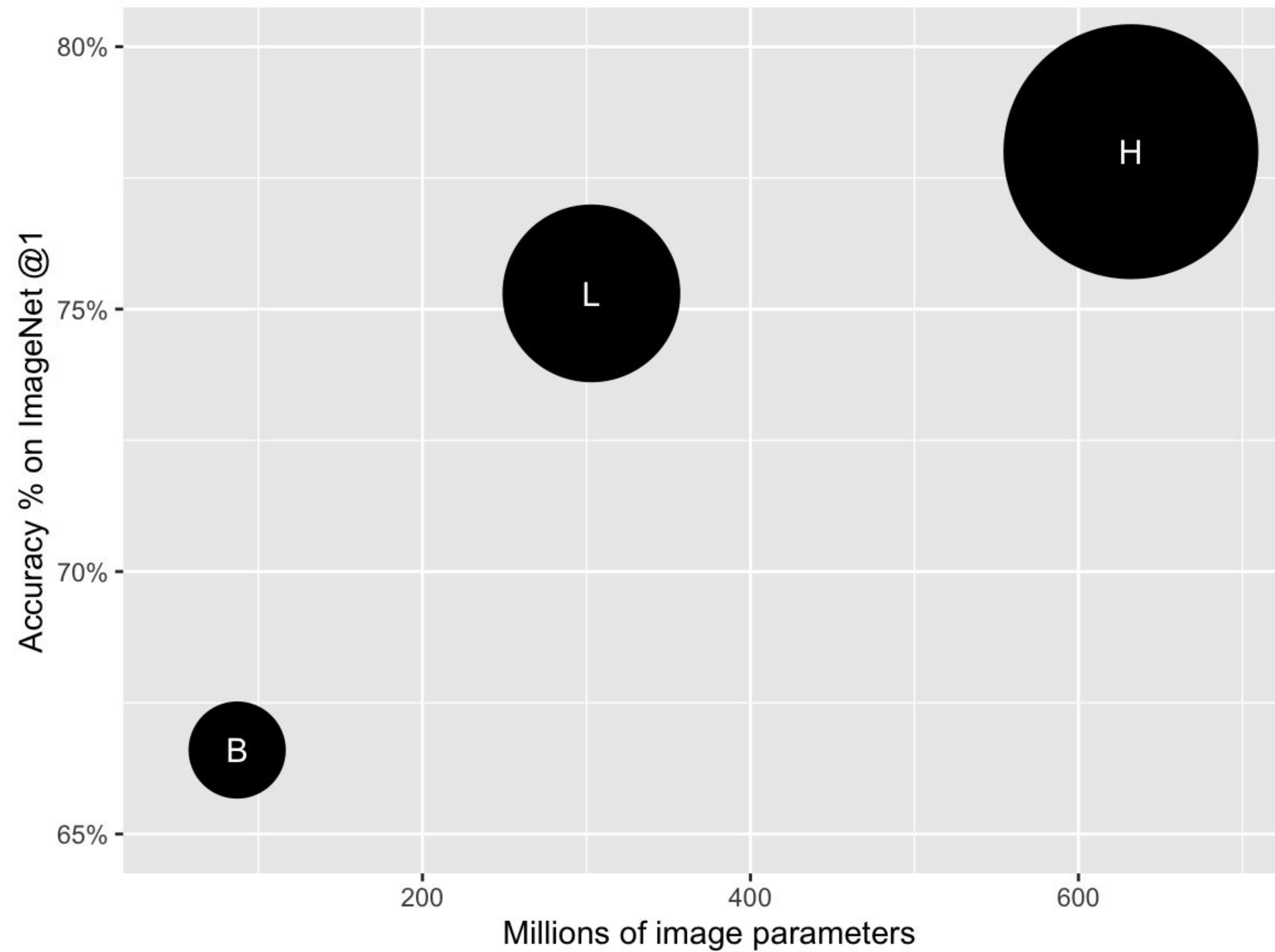
| Model                                                | Training data | Resolution | # of samples seen | ImageNet zero-shot acc. |
|------------------------------------------------------|---------------|------------|-------------------|-------------------------|
| ConvNext-Base                                        | LAION-2B      | 256px      | 13B               | 71.5%                   |
| ConvNext-Large                                       | LAION-2B      | 320px      | 29B               | 76.9%                   |
| ConvNext-XXLarge                                     | LAION-2B      | 256px      | 34B               | 79.5%                   |
| ViT-B/32                                             | DataComp-1B   | 256px      | 34B               | 72.8%                   |
| ViT-B/16                                             | DataComp-1B   | 224px      | 13B               | 73.5%                   |
| ViT-L/14                                             | LAION-2B      | 224px      | 32B               | 75.3%                   |
| ViT-H/14                                             | LAION-2B      | 224px      | 32B               | 78.0%                   |
| ViT-L/14                                             | DataComp-1B   | 224px      | 13B               | 79.2%                   |
| ViT-G/14                                             | LAION-2B      | 224px      | 34B               | 80.1%                   |
|                                                      |               |            |                   |                         |
| ViT-L/14-quickgelu ( <a href="#">Original CLIP</a> ) | WIT           | 224px      | 13B               | 75.5%                   |
| ViT-SO400M/14 ( <a href="#">SigLIP</a> )             | WebLI         | 224px      | 45B               | 82.0%                   |
| ViT-L/14 ( <a href="#">DFN</a> )                     | DFN-2B        | 224px      | 39B               | 82.2%                   |
| ViT-SO400M-14-SigLIP-384 ( <a href="#">SigLIP</a> )  | WebLI         | 384px      | 45B               | 83.1%                   |
| ViT-H/14-quickgelu ( <a href="#">DFN</a> )           | DFN-5B        | 224px      | 39B               | 83.4%                   |
| ViT-H-14-378-quickgelu ( <a href="#">DFN</a> )       | DFN-5B        | 378px      | 44B               | 84.4%                   |



# openCLIP

## Model sizes and accuracy

CLIP size vs accuracy



# openCLIP

DataComp shows bigger data is not always better

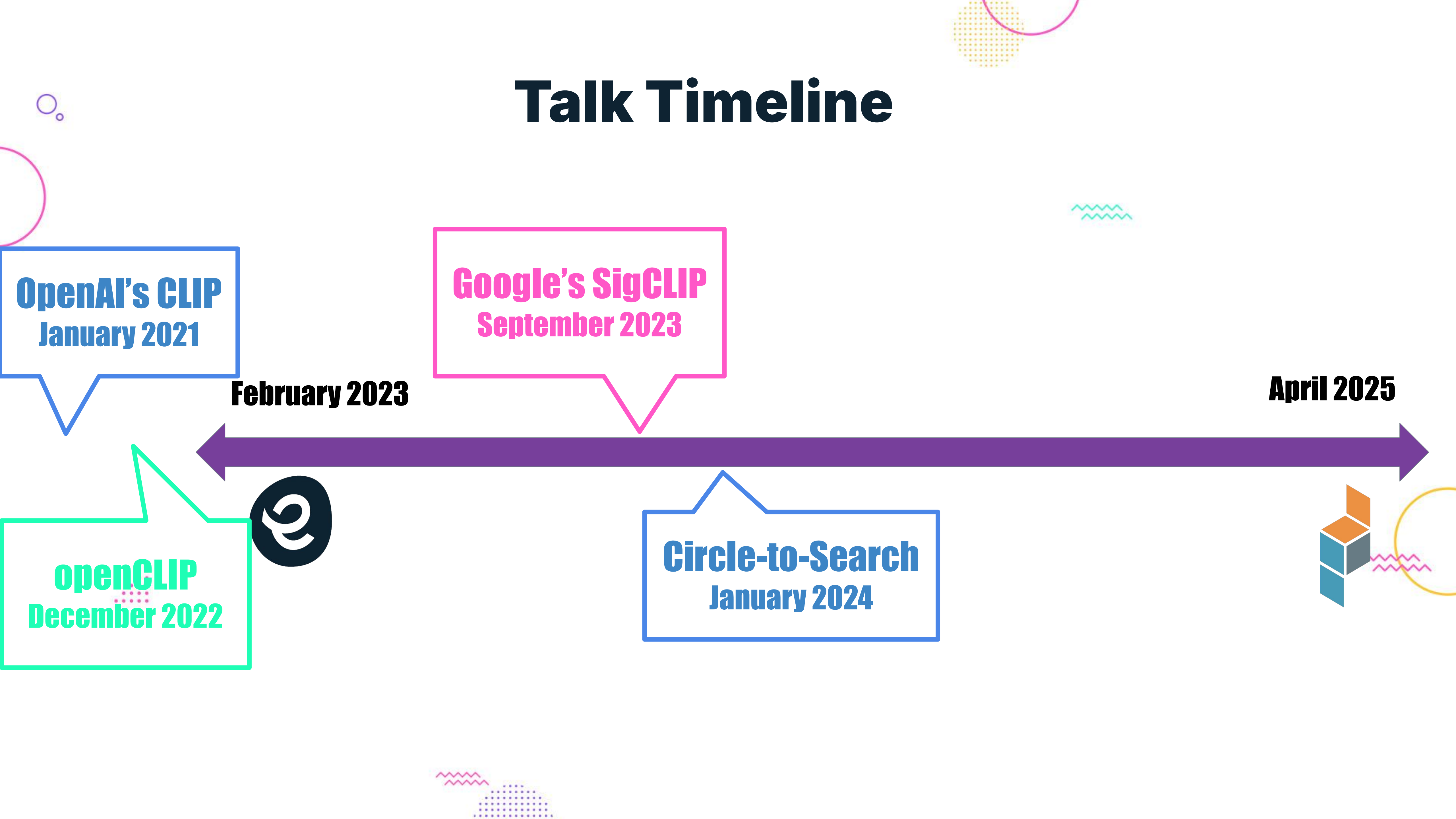
| Model                                                | Training data | Resolution | # of samples seen | ImageNet zero-shot acc. |
|------------------------------------------------------|---------------|------------|-------------------|-------------------------|
| ConvNext-Base                                        | LAION-2B      | 256px      | 13B               | 71.5%                   |
| ConvNext-Large                                       | LAION-2B      | 320px      | 29B               | 76.9%                   |
| ConvNext-XXLarge                                     | LAION-2B      | 256px      | 34B               | 79.5%                   |
| ViT-B/32                                             | DataComp-1B   | 256px      | 34B               | 72.8%                   |
| ViT-B/16                                             | DataComp-1B   | 224px      | 13B               | 73.5%                   |
| ViT-L/14                                             | LAION-2B      | 224px      | 32B               | 75.3%                   |
| ViT-H/14                                             | LAION-2B      | 224px      | 32B               | 78.0%                   |
| ViT-L/14                                             | DataComp-1B   | 224px      | 13B               | 79.2%                   |
| ViT-G/14                                             | LAION-2B      | 224px      | 34B               | 80.1%                   |
| ViT-L/14-quickgelu ( <a href="#">Original CLIP</a> ) | WIT           | 224px      | 13B               | 75.5%                   |
| ViT-SO400M/14 ( <a href="#">SigLIP</a> )             | WebLI         | 224px      | 45B               | 82.0%                   |
| ViT-L/14 ( <a href="#">DFN</a> )                     | DFN-2B        | 224px      | 39B               | 82.2%                   |
| ViT-SO400M-14-SigLIP-384 ( <a href="#">SigLIP</a> )  | WebLI         | 384px      | 45B               | 83.1%                   |
| ViT-H/14-quickgelu ( <a href="#">DFN</a> )           | DFN-5B        | 224px      | 39B               | 83.4%                   |
| ViT-H-14-378-quickgelu ( <a href="#">DFN</a> )       | DFN-5B        | 378px      | 44B               | 84.4%                   |

**openCLIP**  
December 2022

**openCLIP**  
**w/ DataComp**  
April 2023



# Talk Timeline



Source Input:

# 1) Reverse Image Search

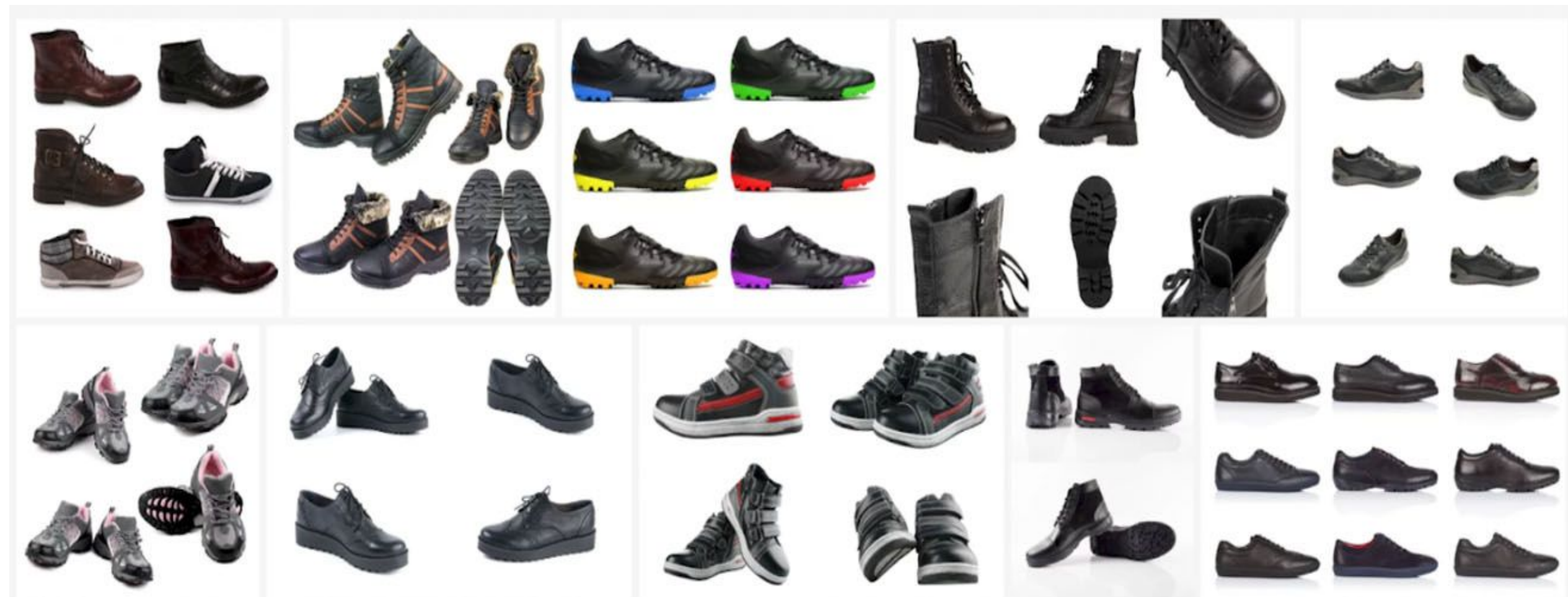
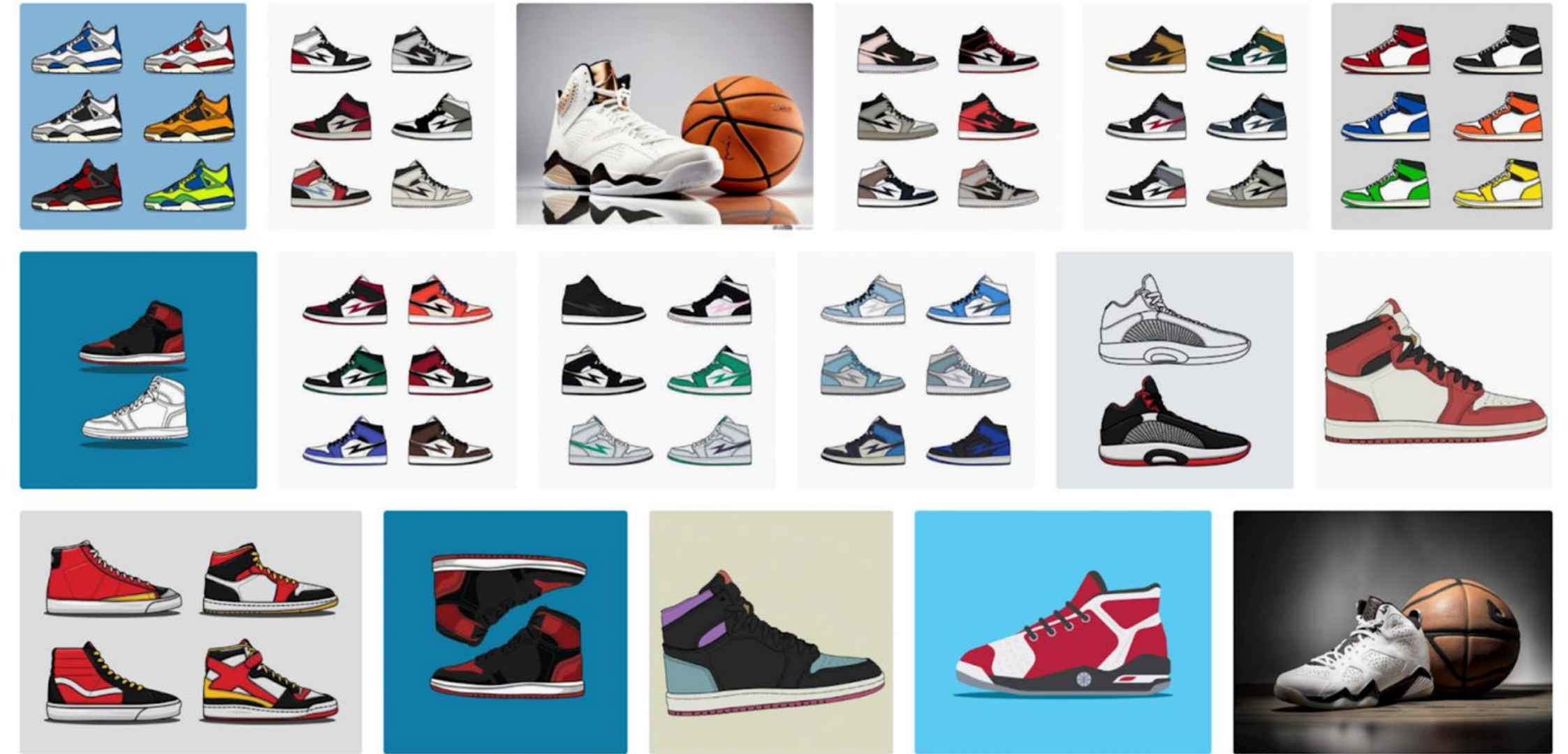




Source Input:



# 1) Reverse Image Search

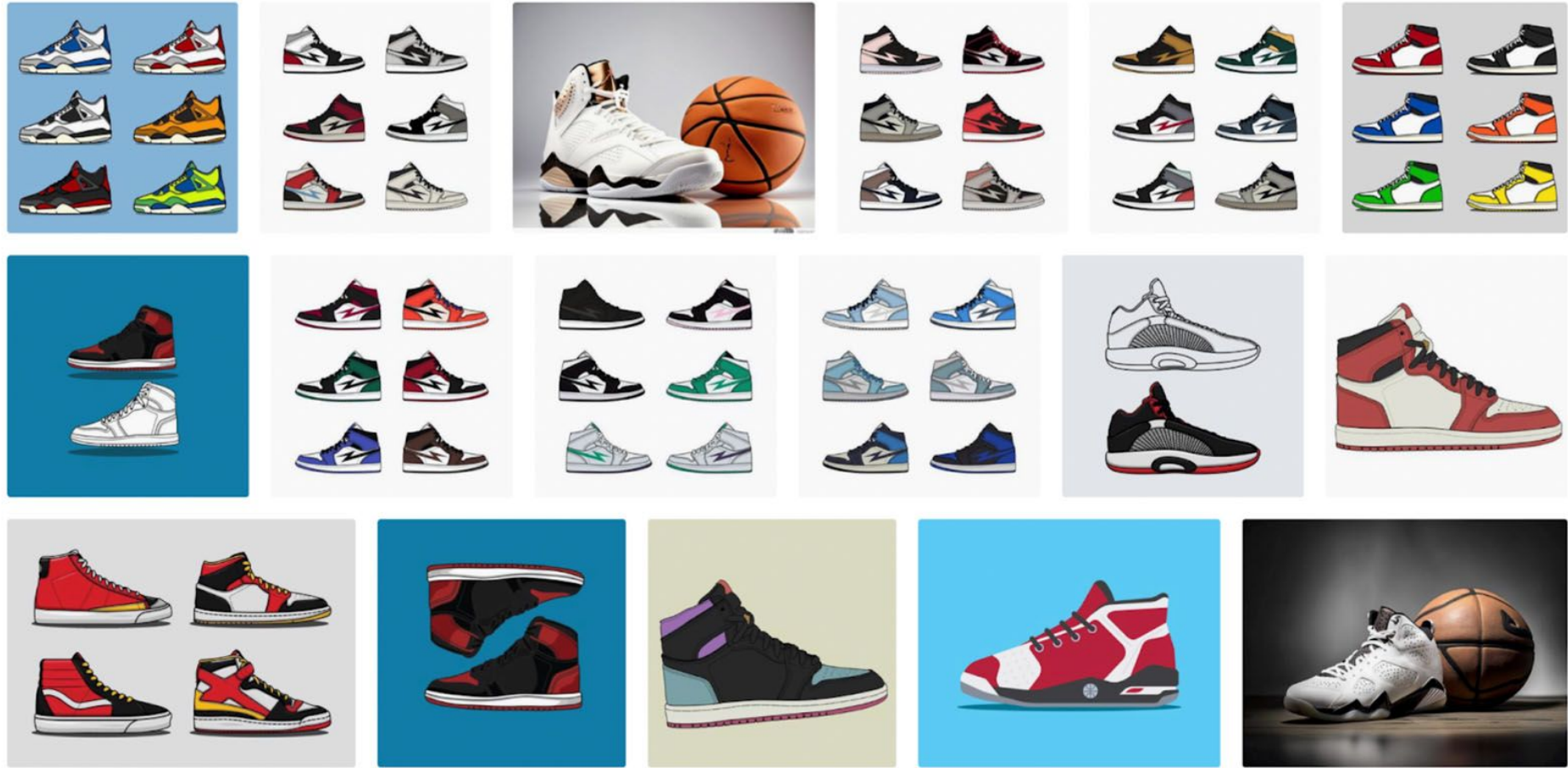


NOT 

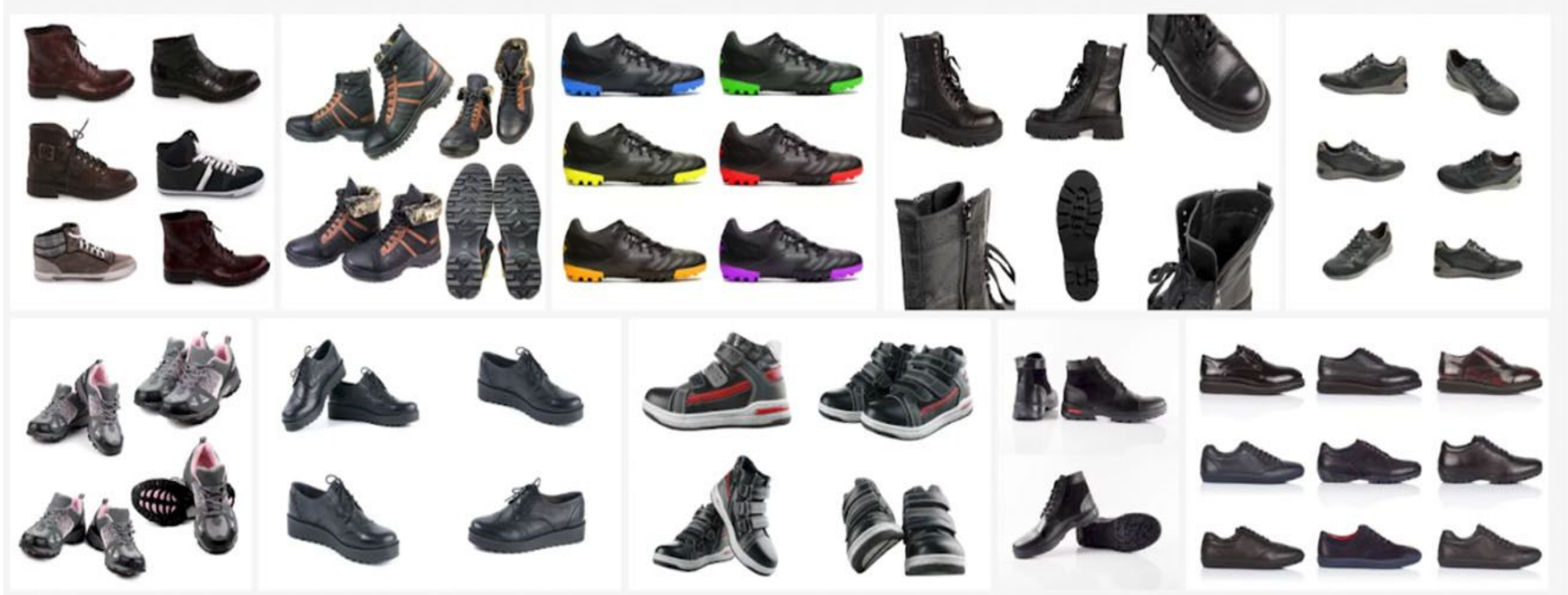


Source Input:

# 1) Reverse Image Search



● AB-test Results:



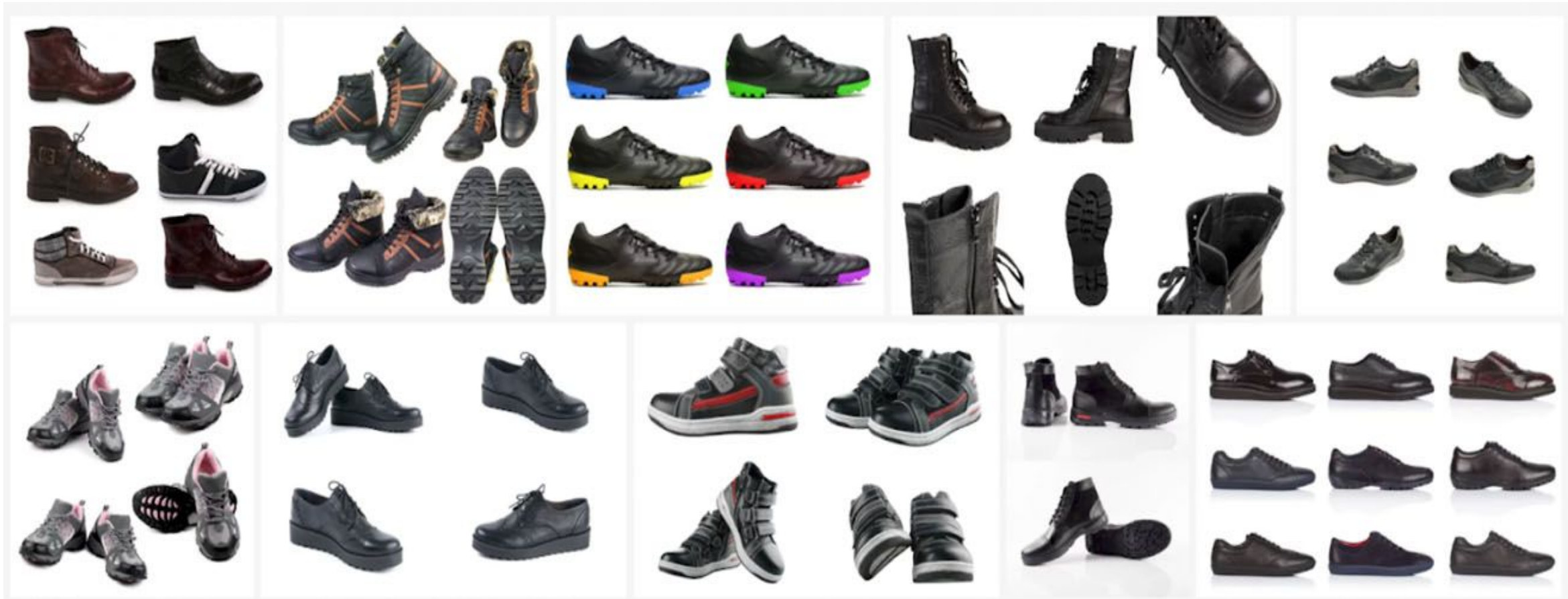
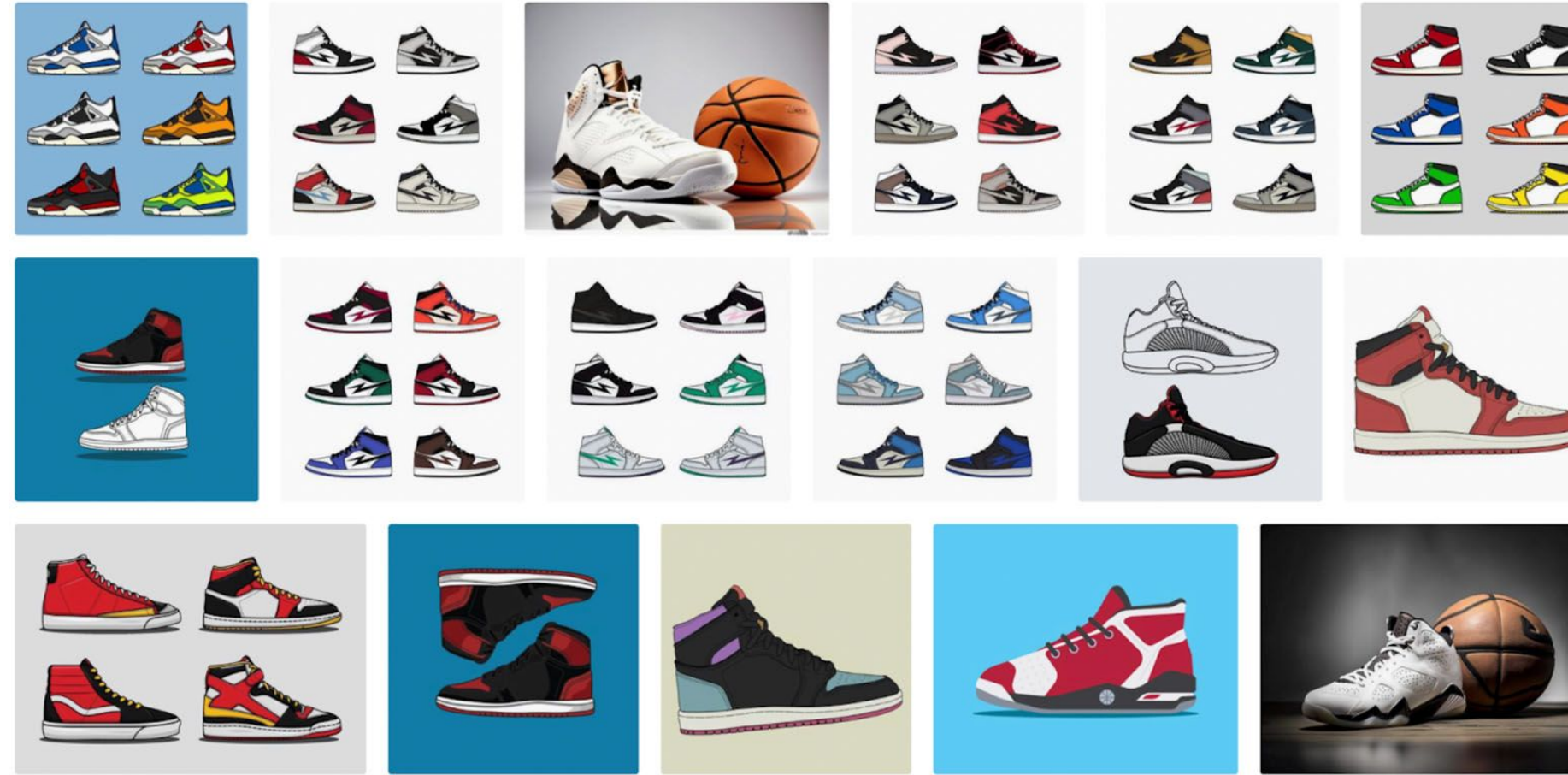
NOT





Source Input:

# 1) Reverse Image Search



NOT





# Talk Timeline

February 2023

April 2025

Reverse Image Search  
May 2023





## 2) Search Bar

Legacy Algo

### User Query

["background", "crumpled paper", "easter card"]



## 2) Search Bar

Legacy Algo

**ConstantScore**  
(capped token matching)



## 2) Search Bar

Legacy Algo

**BehavioralScore**  
(recent interactions)

**ConstantScore**  
(capped token matching)



## 2) Search Bar

Legacy Algo

**TieBreakers**  
(aesthetics, freshness)

**BehavioralScore**  
(recent interactions)

**ConstantScore**  
(capped token matching)



## 2) Search Bar

Legacy Algo

**TieBreakers**

(aesthetics freshness)



# elasticsearch

**ConstantScore**

(capped token matching)



## 2) Search Bar

First algo

### User Query

["background", "crumpled paper", "easter card"]



## 2) Search Bar

First algo

**Tie Breakers**

**Behavioral Score**

**Constant Score**

**CLIP Score**



## 2) Search Bar

Hybrid search stack

Tie Breakers



Constant\* Score

CLIP\* Score



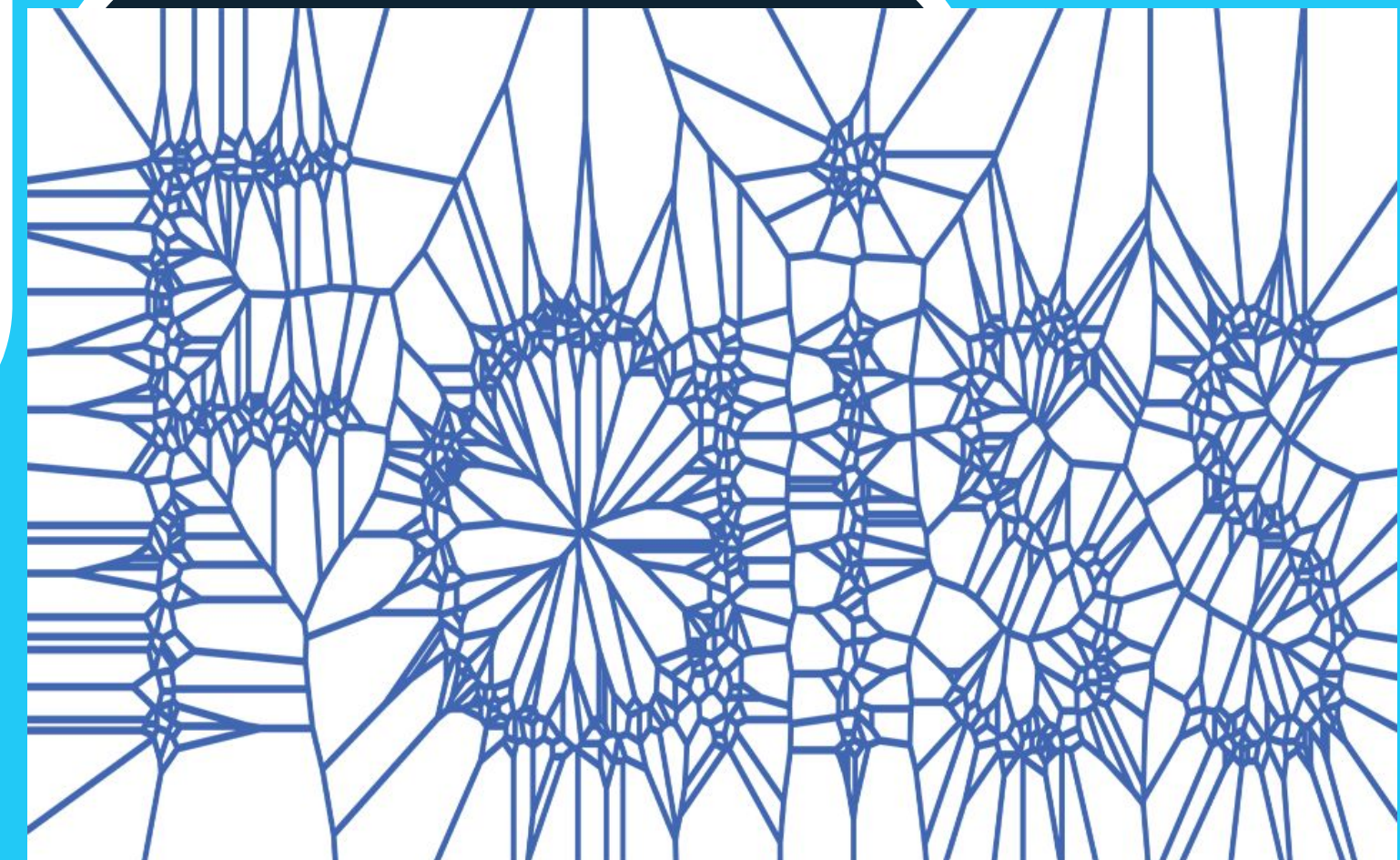
## 2) Search Bar

Hybrid search stack

Tie Breakers



Constant\* Score





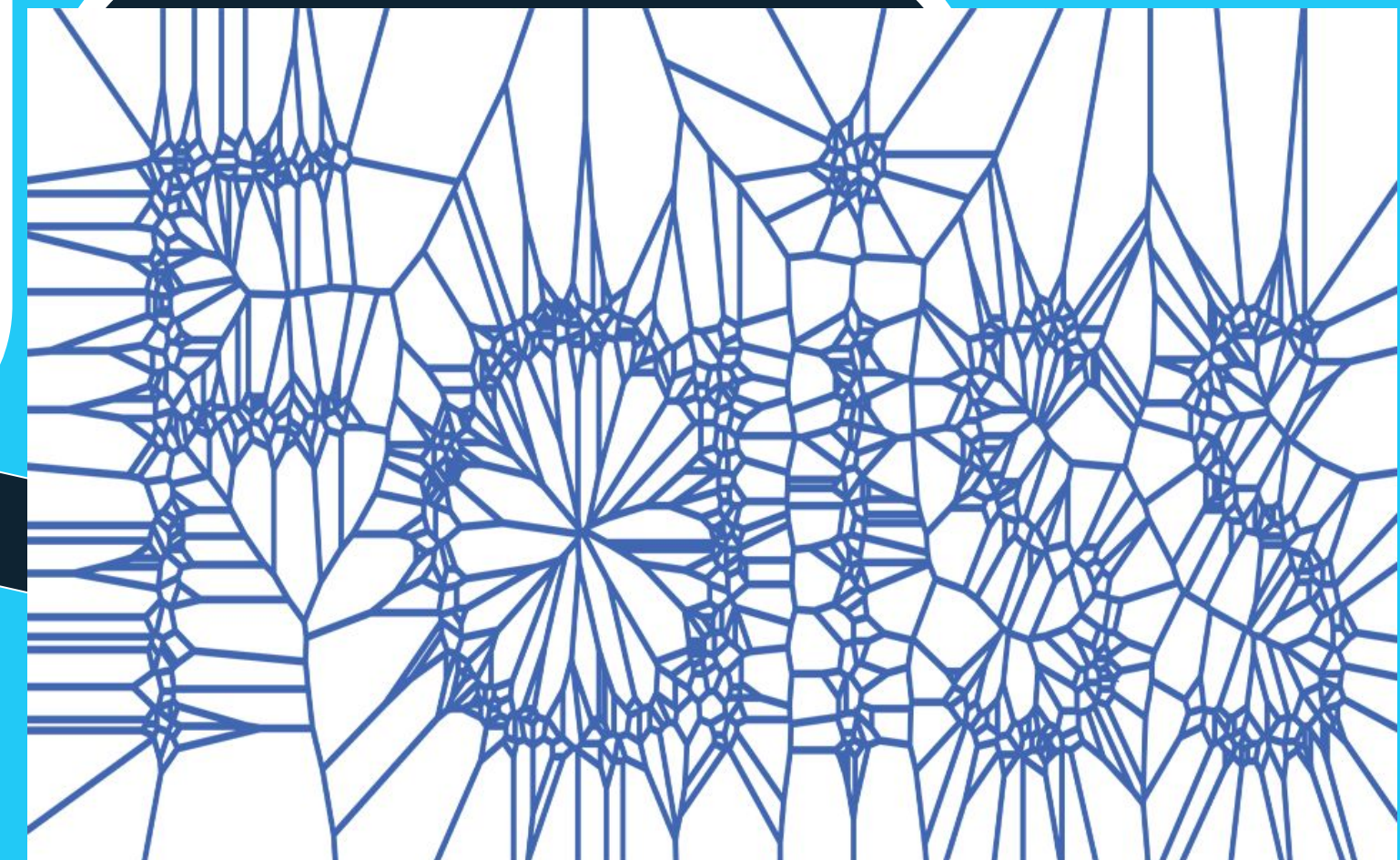
## 2) Search Bar

Hybrid search stack

Tie Breakers



Constant\* Score





## 2) Search Bar

First AB

Downloads  
-8.52%

B  
oral\* S  
(cap  
ken match

Search  
pages  
+2.18%

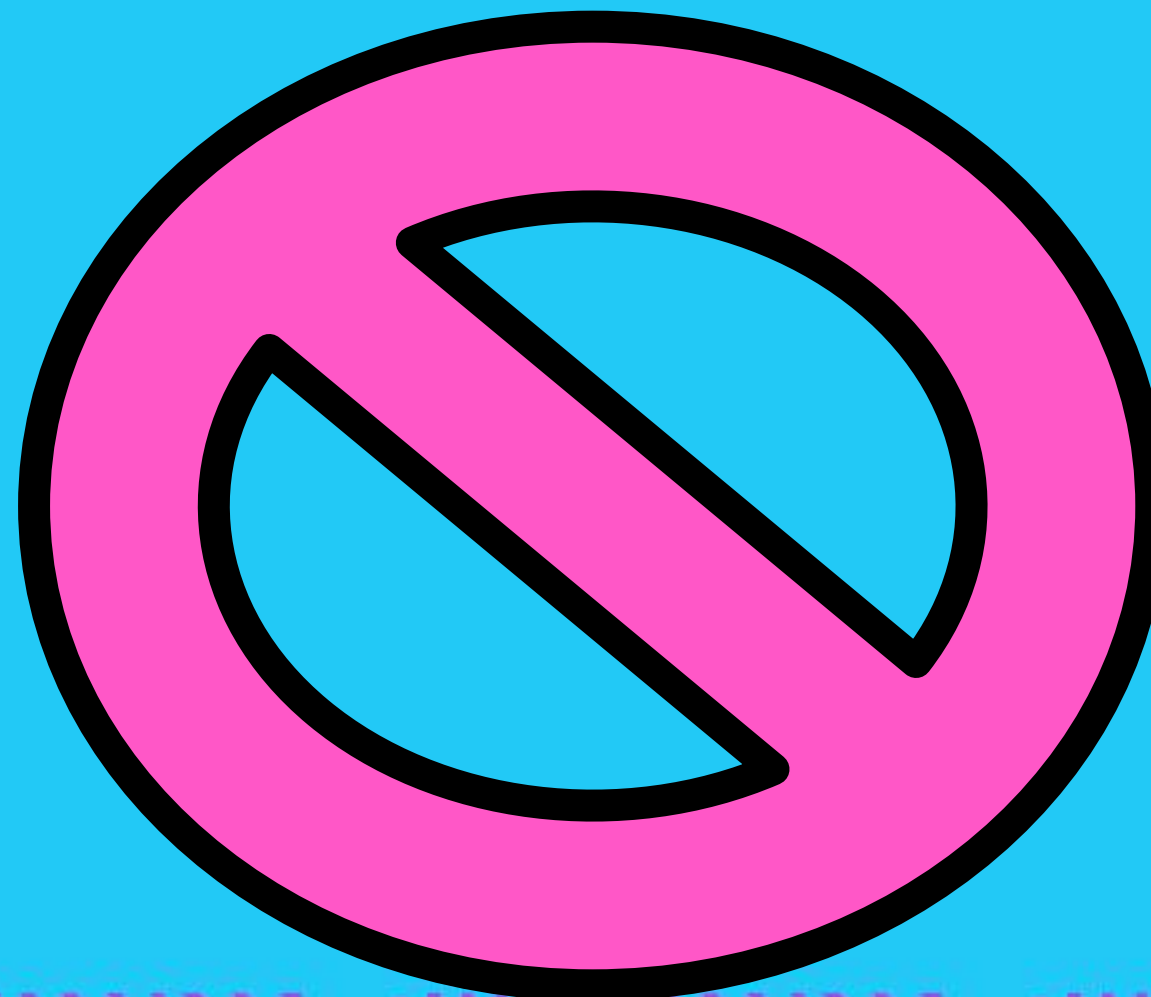
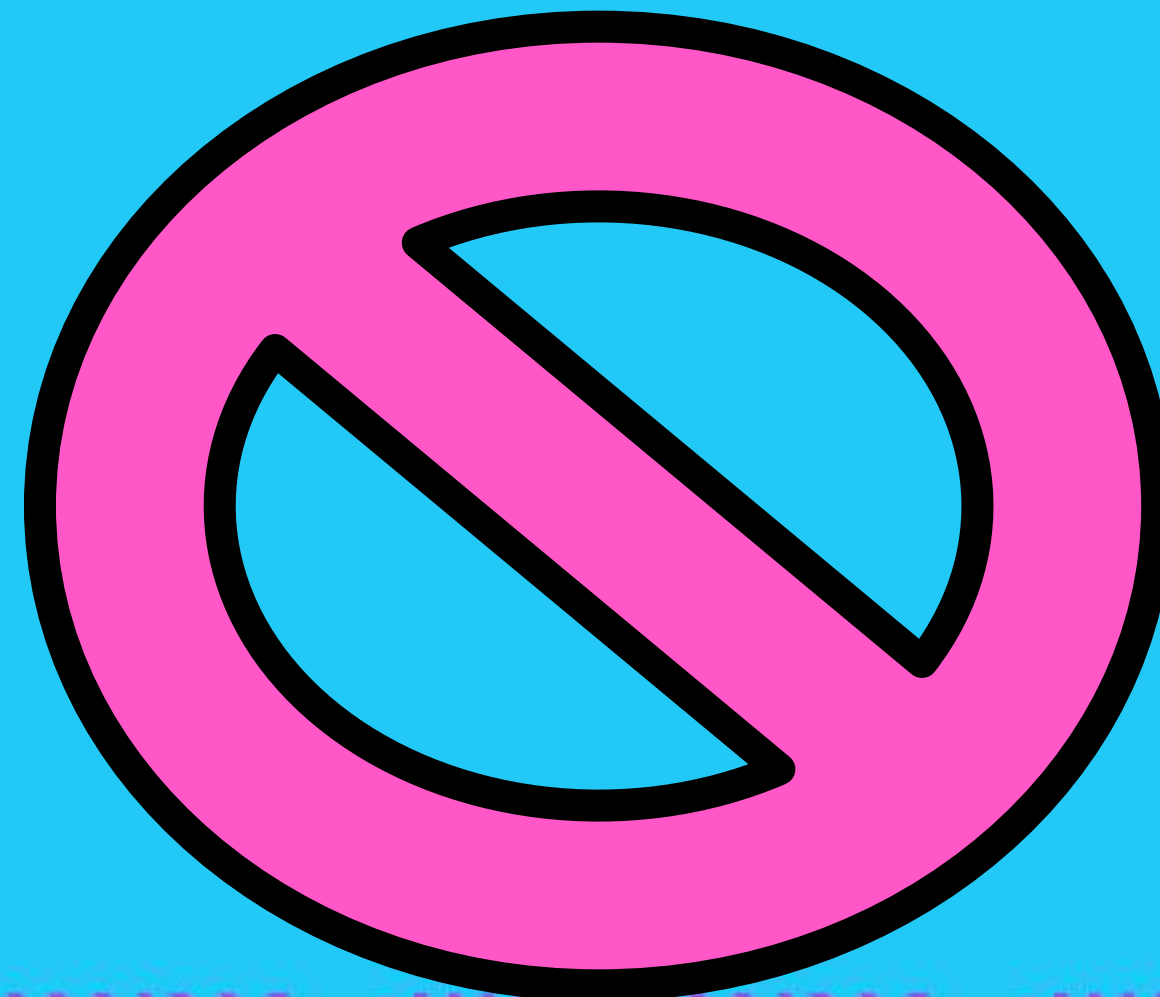
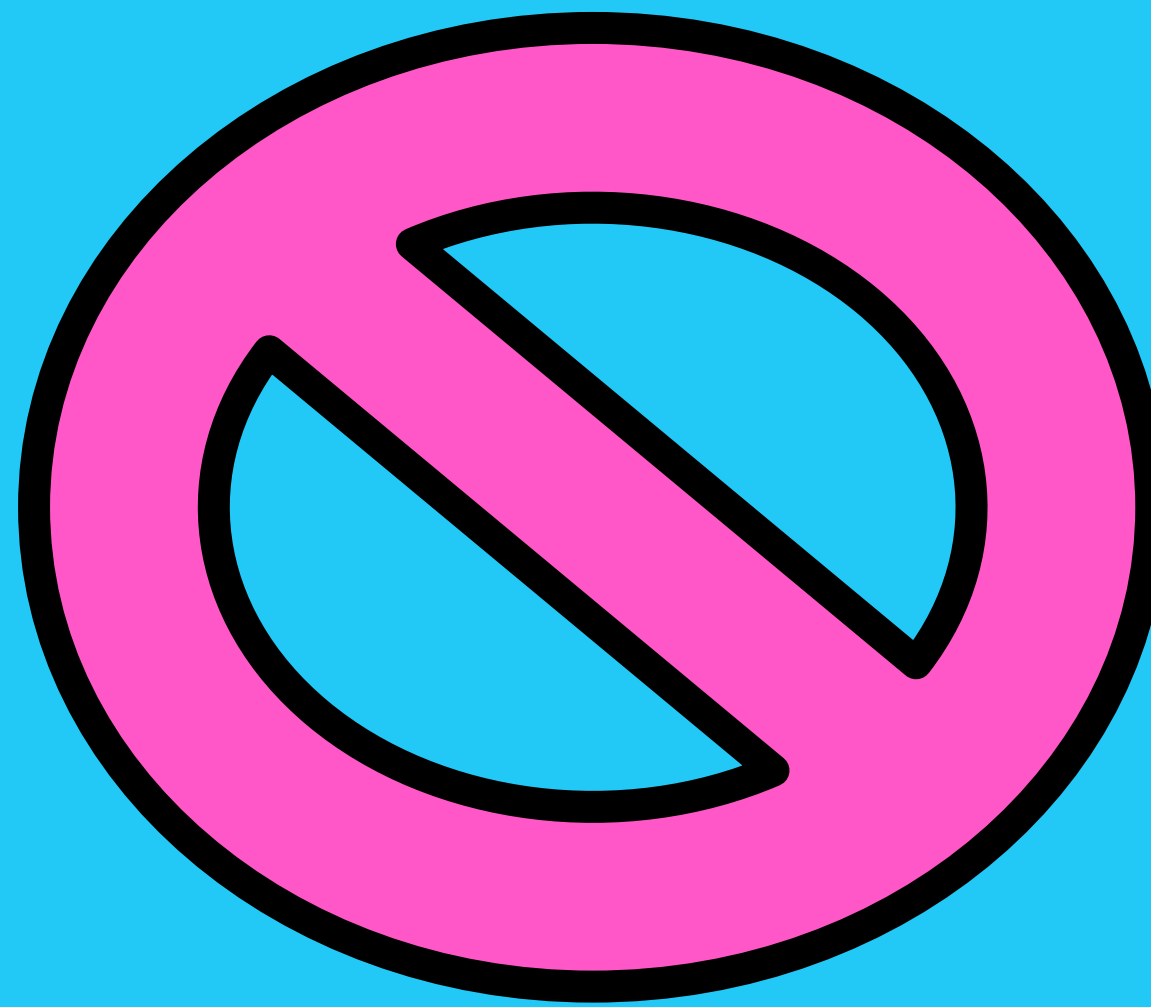
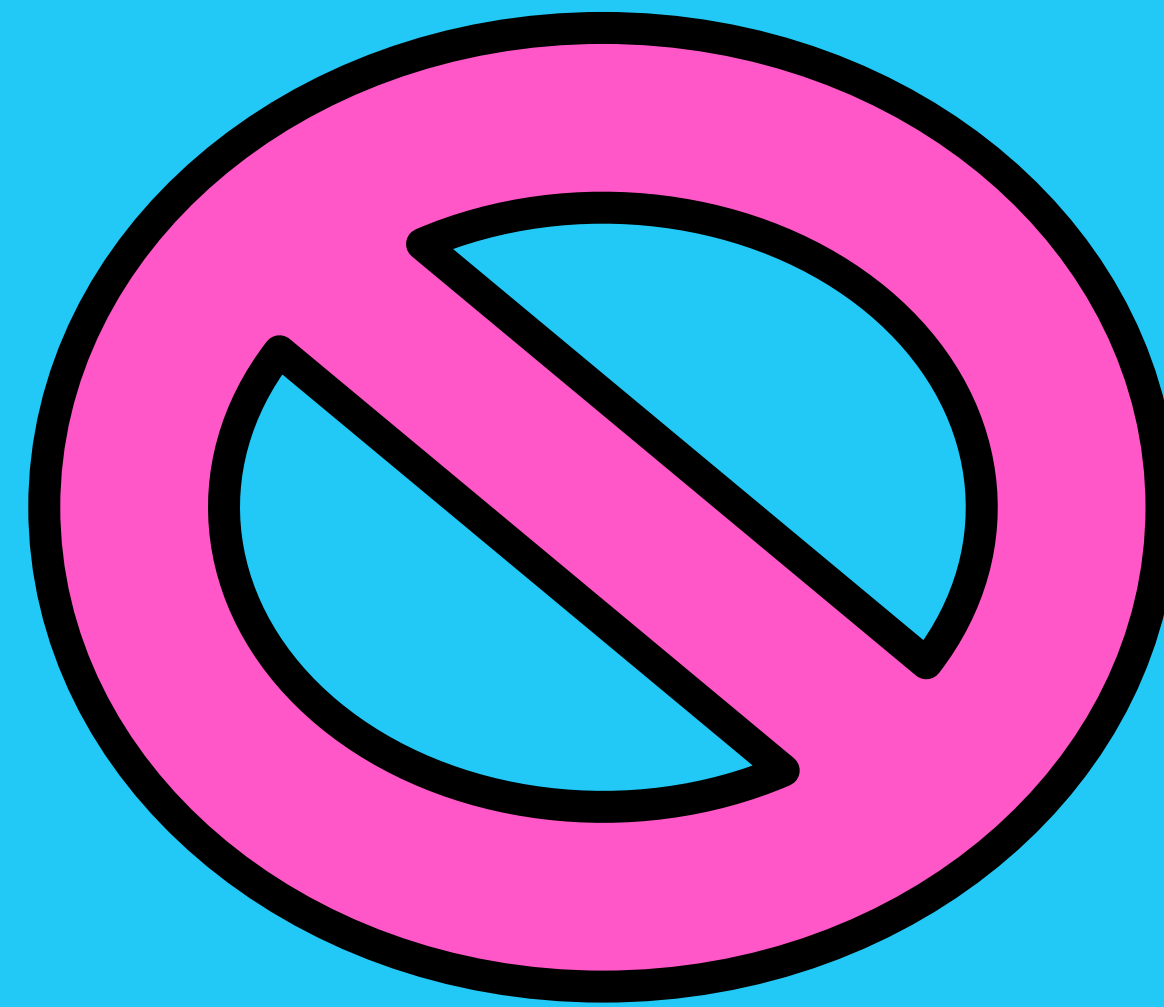
Const

ore



## 2) Search Bar

Next 4 ABs/algos





**Disjunction**  
**Maximum**

## 2) Search Bar

Sixth (and final) algo

Tie Breakers

Behavioral Score

CLIP Score

Constant Score

[“k

”]



**Disjunction**  
**Maximum**

## 2) Search Bar

Sixth (and final) algo

Tie Breakers

Behavioral Score

CLIP Score

Constant Score





## 2) Search Bar

Final AB

Searches  
+3.28%

Revenue  
+12.46%

Behavior



CLIP Score

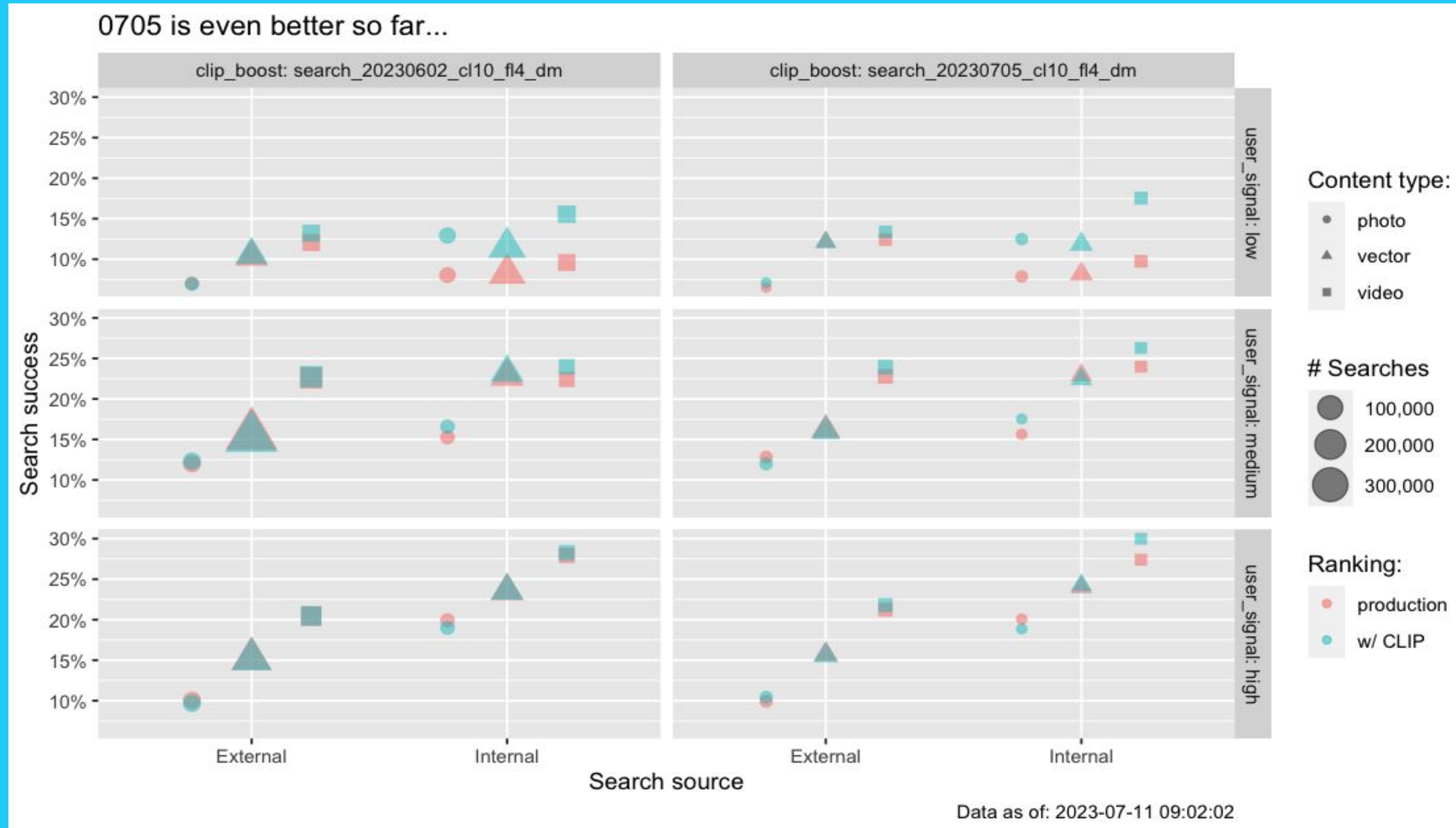
Constant Score

Downloads  
+2.02%



## 2) Search Bar

Final algo





## 2) Search Bar

**Final Algo**  
**mid-July 2023**





## 2) Search Bar





# Talk Timeline

February 2023

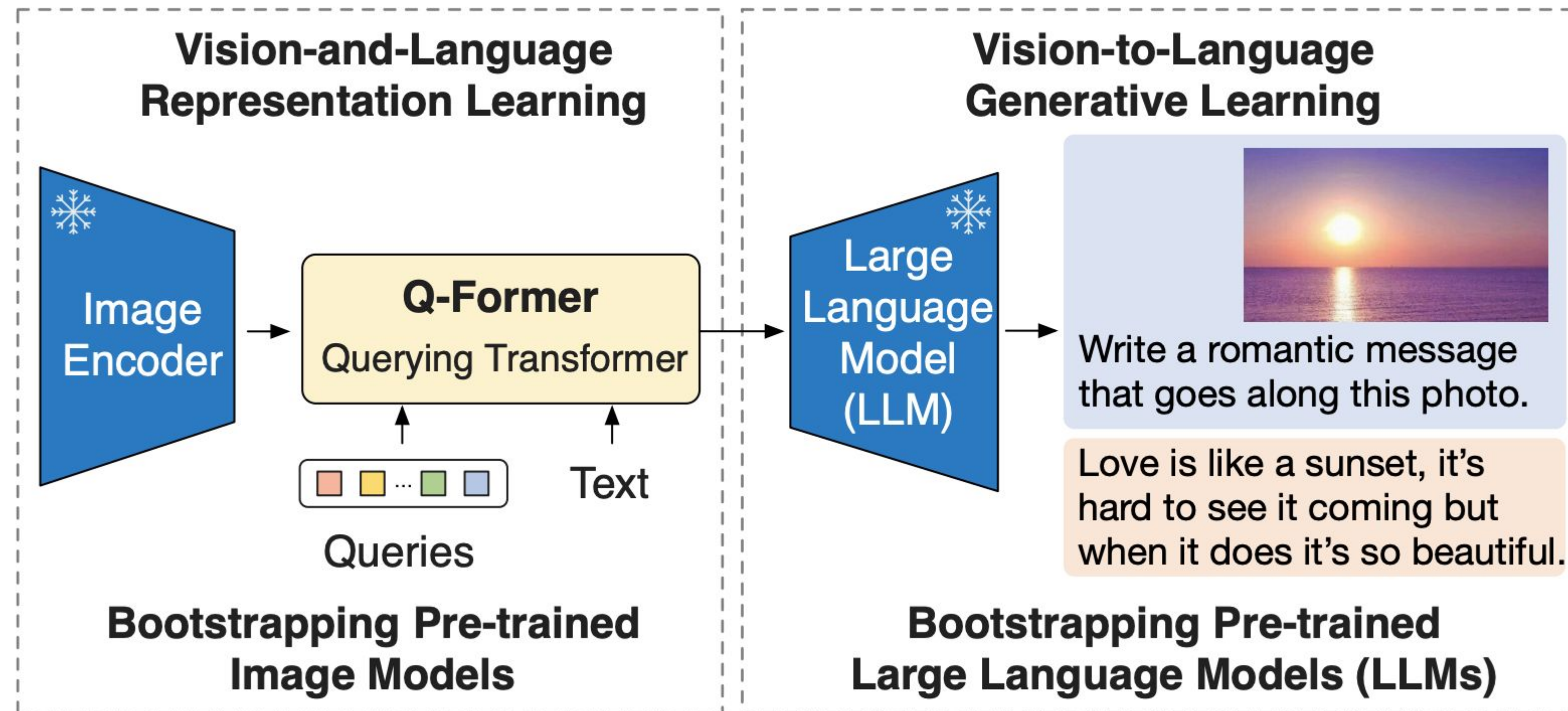
Search Bar  
July 2023

April 2025



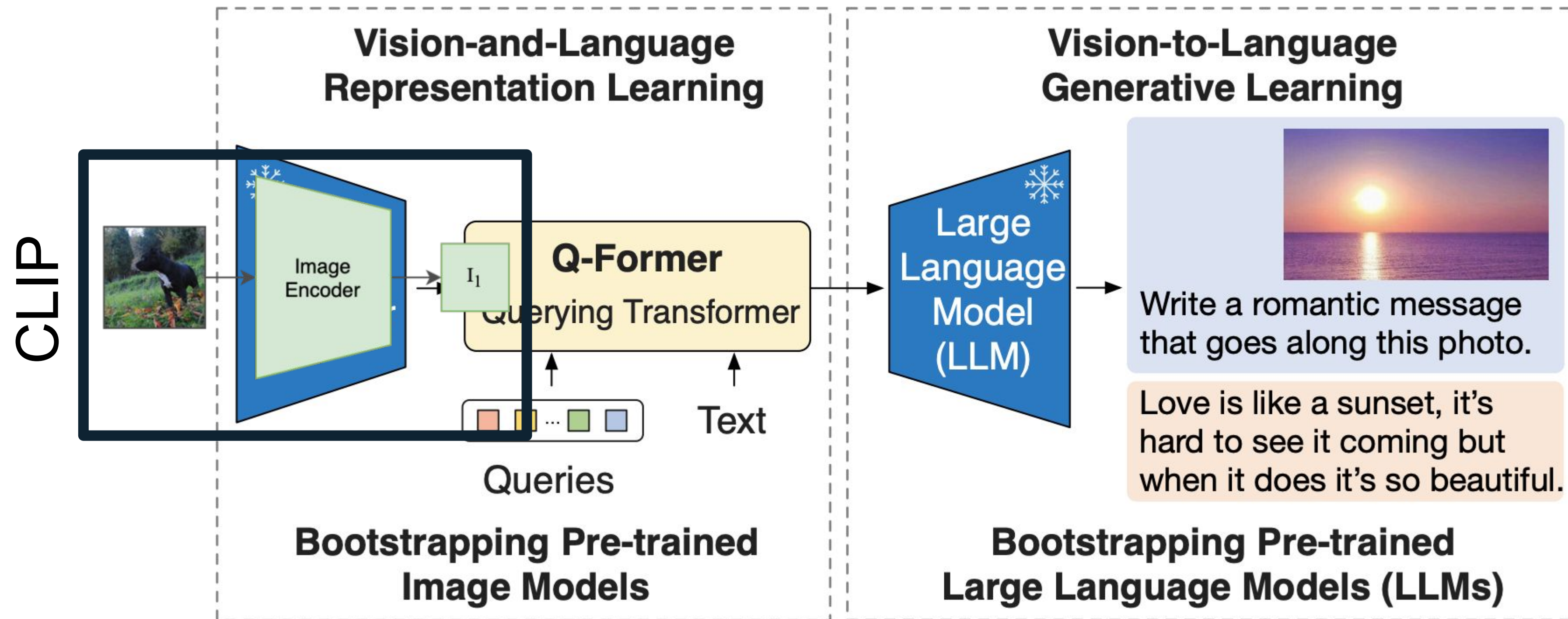


### 3) BLIP





### 3) BLIP





# 3) Magic Metadata



DashboardMy Portfolio ▼Distribution BETA ▼Insights ▼ReferralsEarnings ▼Support

UploadAdd Data (99)Pending ReviewApprovedNot Approved

Content Type: All ▼License Type: All ▼Sort: Newest ▼

Select all

CSV

More



✓ Rain-streaked window captures blurred ... P AI



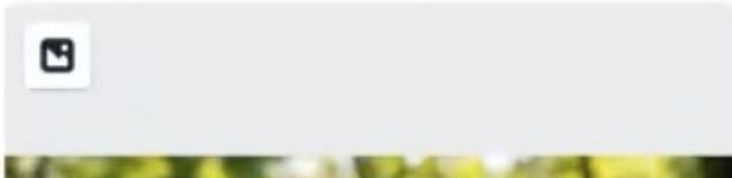
✓ A Young Man Praying Under The Light I... P AI



✓ Succulent Cooked Lobster with Lemon ... P AI



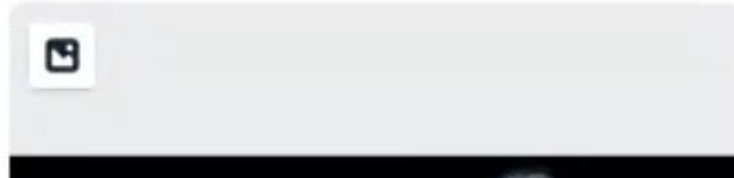
✓ A mesmerizing dance of vibrant orange ... P AI



✓ Picnic Day with a Couple enjoying sunny... P AI



✓ Succulent Steak with Rich Sauce, Steame... P AI



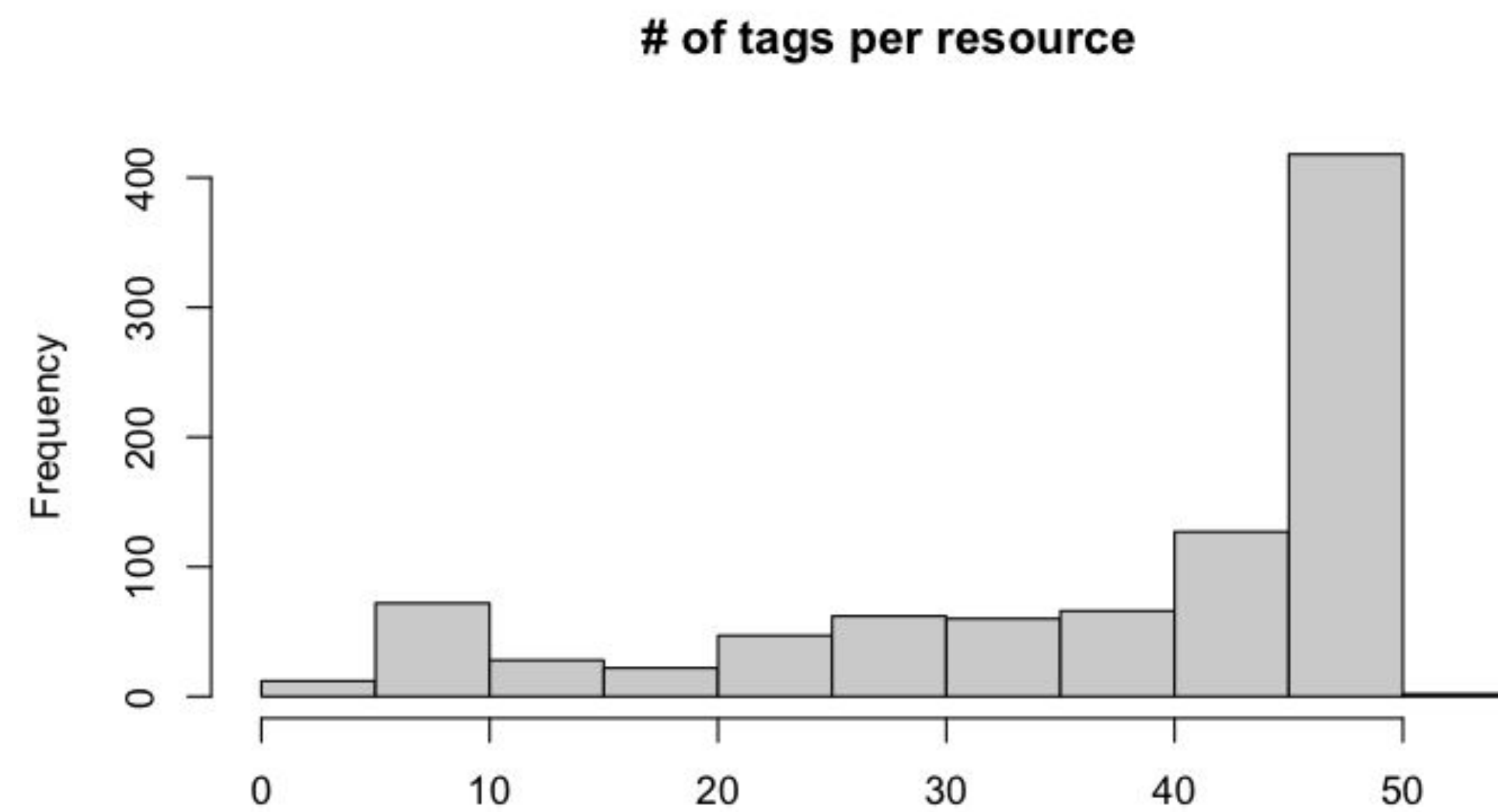
✓ Ethereal wisps of smoke dance against t... P AI



✓ A Healthy Lunchbox with Rice, Spinach, ... P AI



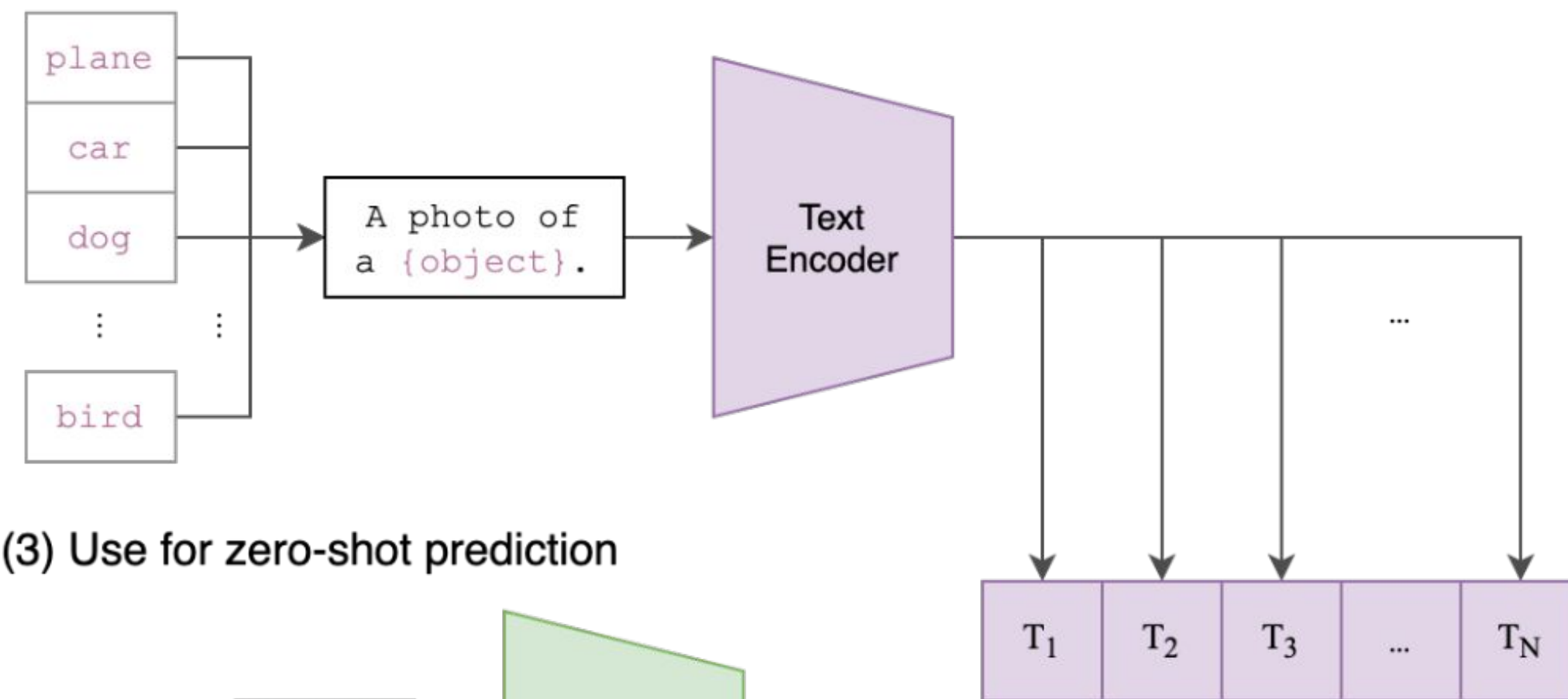
## 3b) Token 🧽



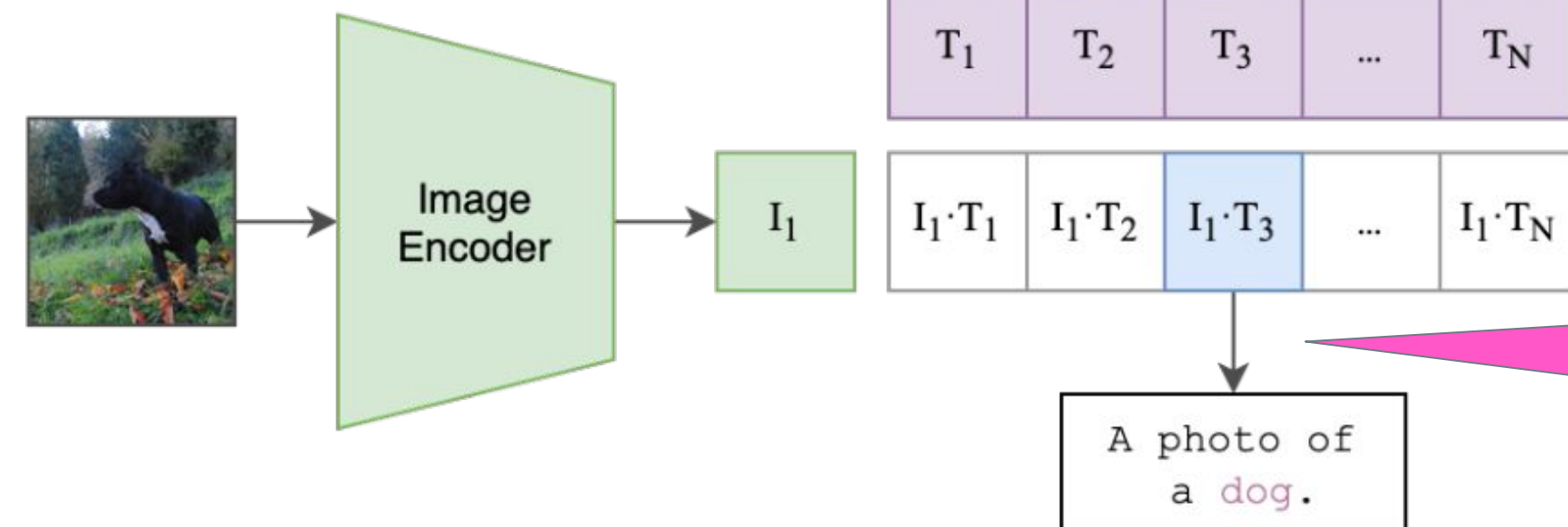


# 3b) Token 🧽

(2) Create dataset classifier from label text



(3) Use for zero-shot prediction

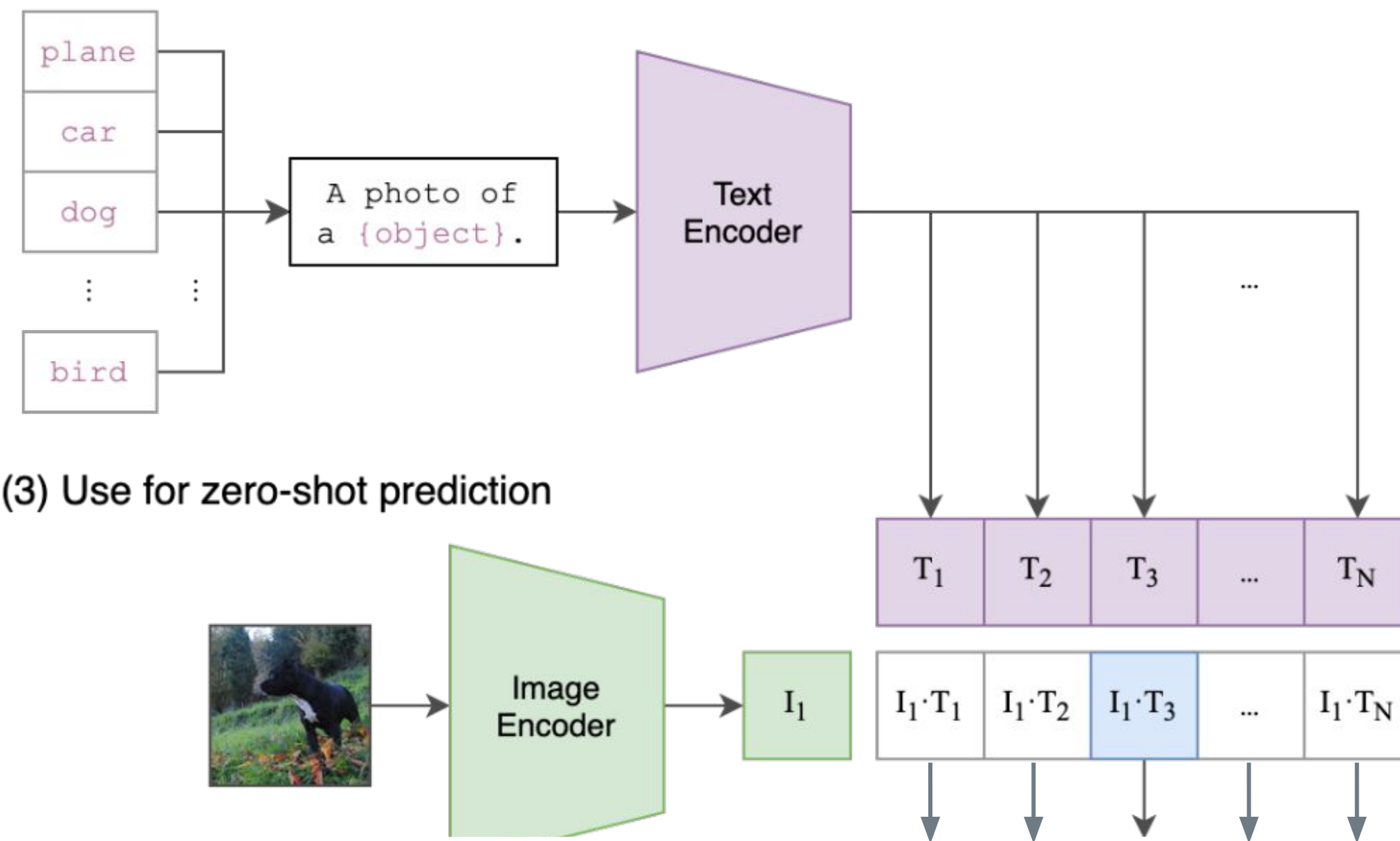


SoftMax



# 3b) Token 🧽

(2) Create dataset classifier from label text



Token  
Similarities





## 3b) Token 🧽

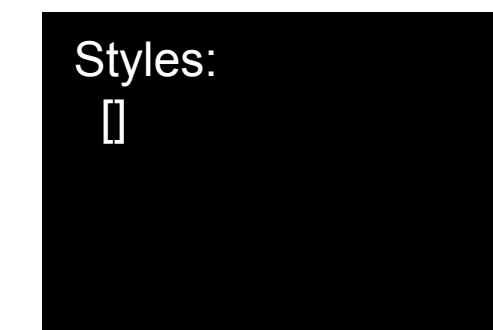
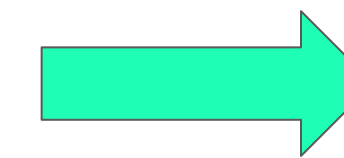
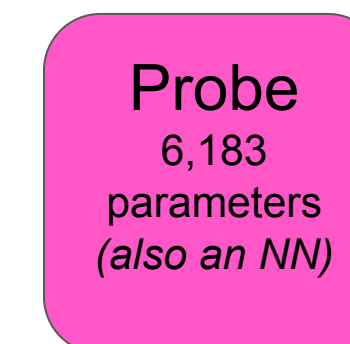
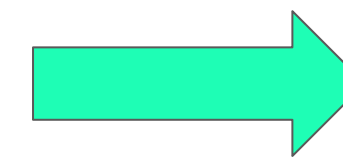
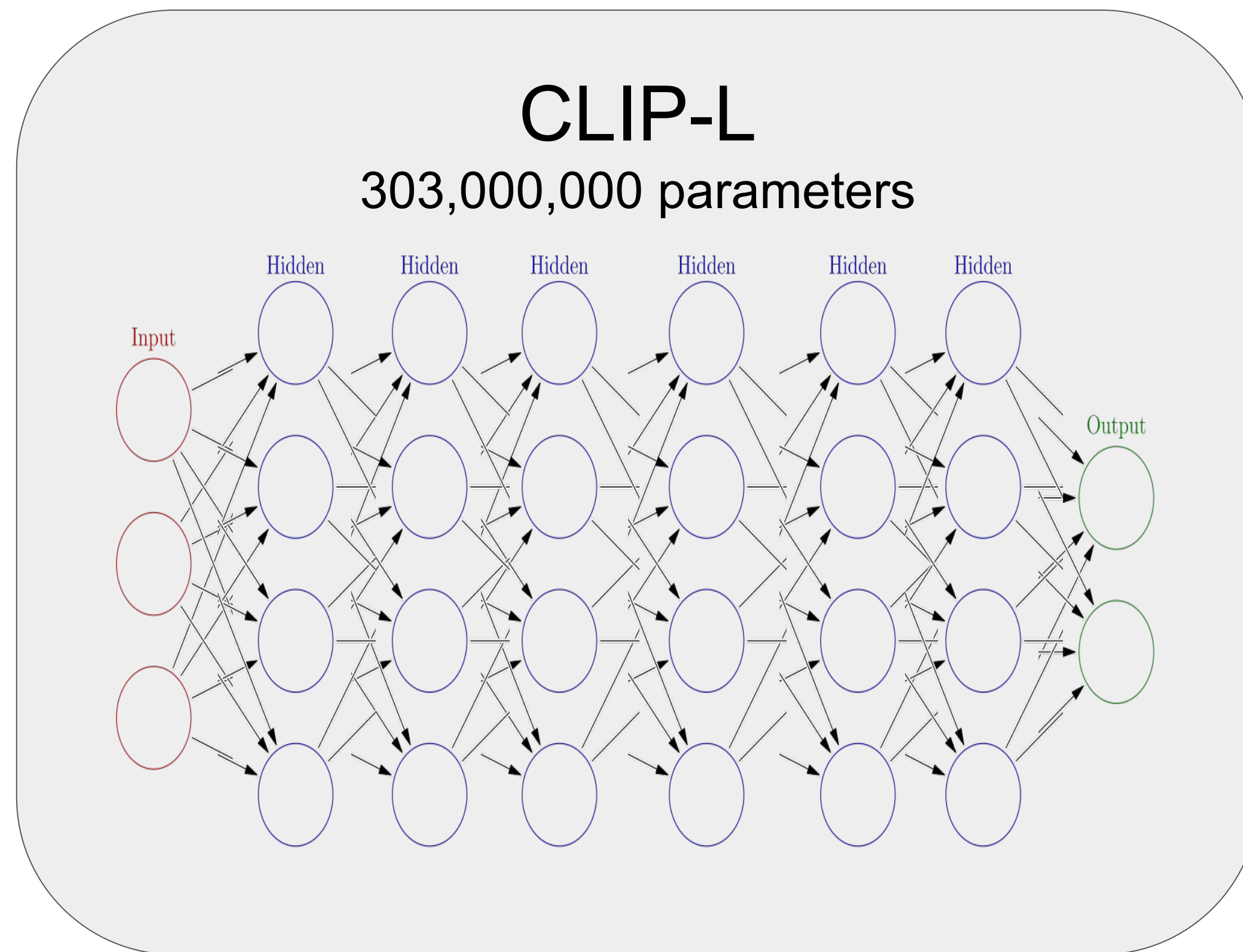
```
[  
('cat', 0.27586338),  
('kitty', 0.25679928),  
('glasses', 0.23471236),  
('pet', 0.23186983),  
('animal', 0.22301413),  
('look', 0.22210719),  
('beautiful', 0.21266793),  
('muzzle', 0.20172776),  
('funny', 0.19799295),  
('lying', 0.17617974),  
('eyes', 0.16986102),  
('bengali', 0.16862199),  
('sofa', 0.16564193),  
('thoroughbred', 0.12821756)  
]
```



```
[  
('cat', 0.27586338),  
('kitty', 0.25679928),  
('glasses', 0.23471236),  
('pet', 0.23186983),  
('animal', 0.22301413),  
('look', 0.22210719),  
('beautiful', 0.21266793)  
]
```



# 3c) Style 🔍





## 3c) Style 🔍



Probe  
6,183  
parameters  
(also an NN)



Styles:  
[  
3D,  
geometric  
]



# 3c) Style



gradient  
geometric-gradient  
22968689



flat  
flat  
48936179



geometric  
geometric  
22950947



handdrawn  
handdrawn  
20297838



flat  
flat  
16390473



flat  
flat  
39341793



geometric  
geometric  
50666174



geometric  
geometric  
2153510



isometric  
isometric-threeed  
5414837



flat  
flat  
8964537



gradient  
gradient  
13159142



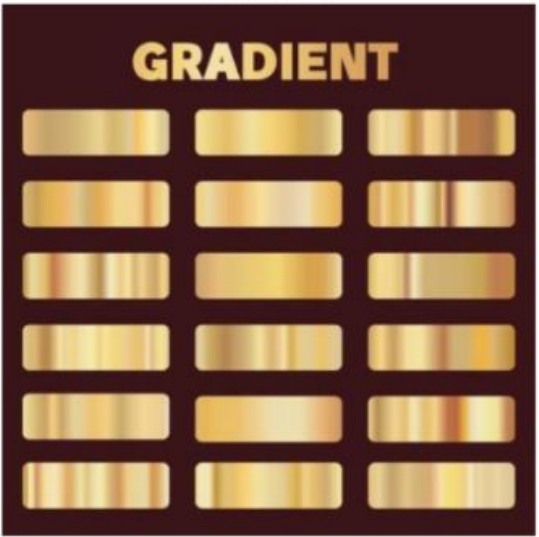
gradient  
gradient  
2964637



threed  
threed  
20929878



gradient  
gradient  
14704854



watercolor  
watercolor  
10725737



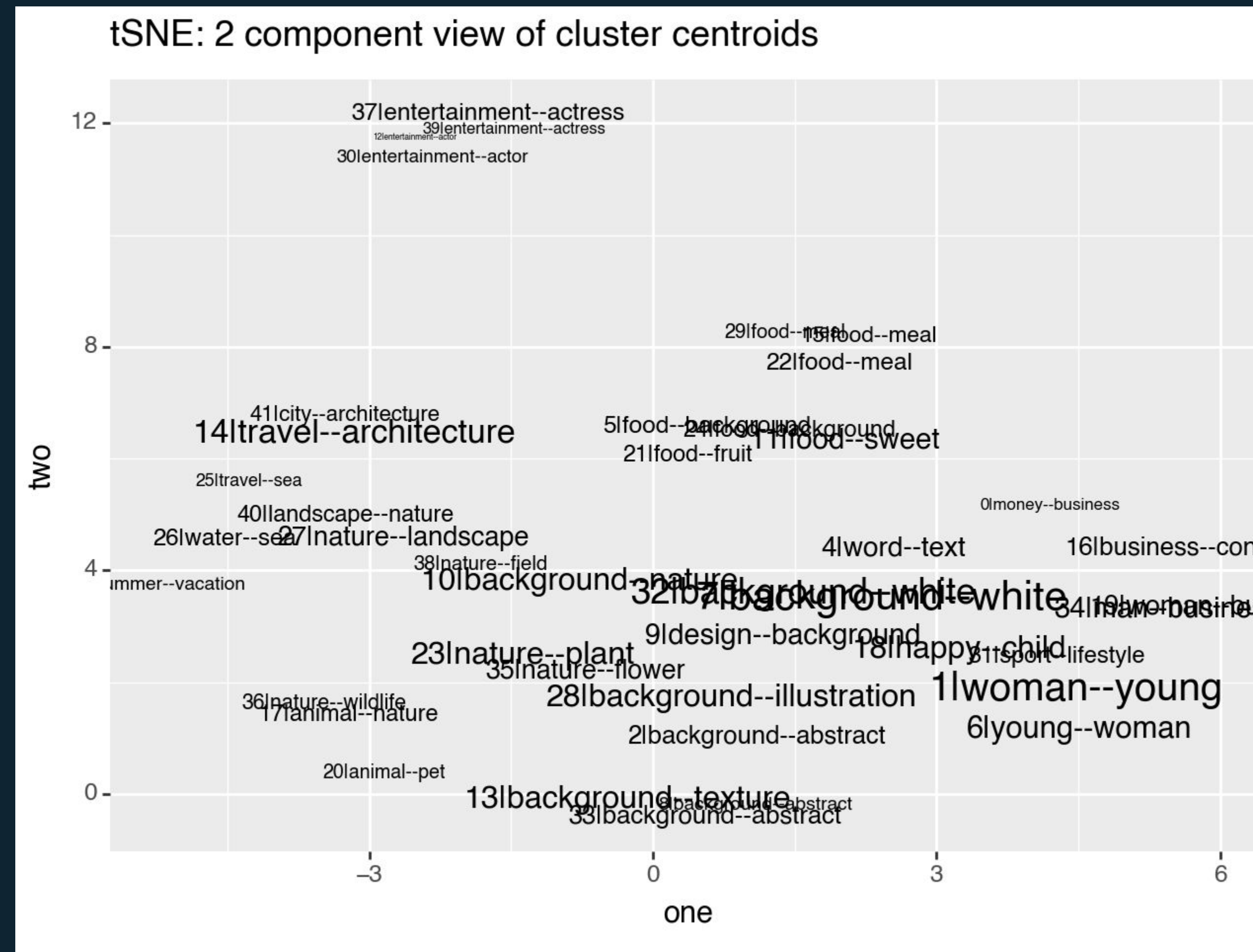


# Talk Timeline



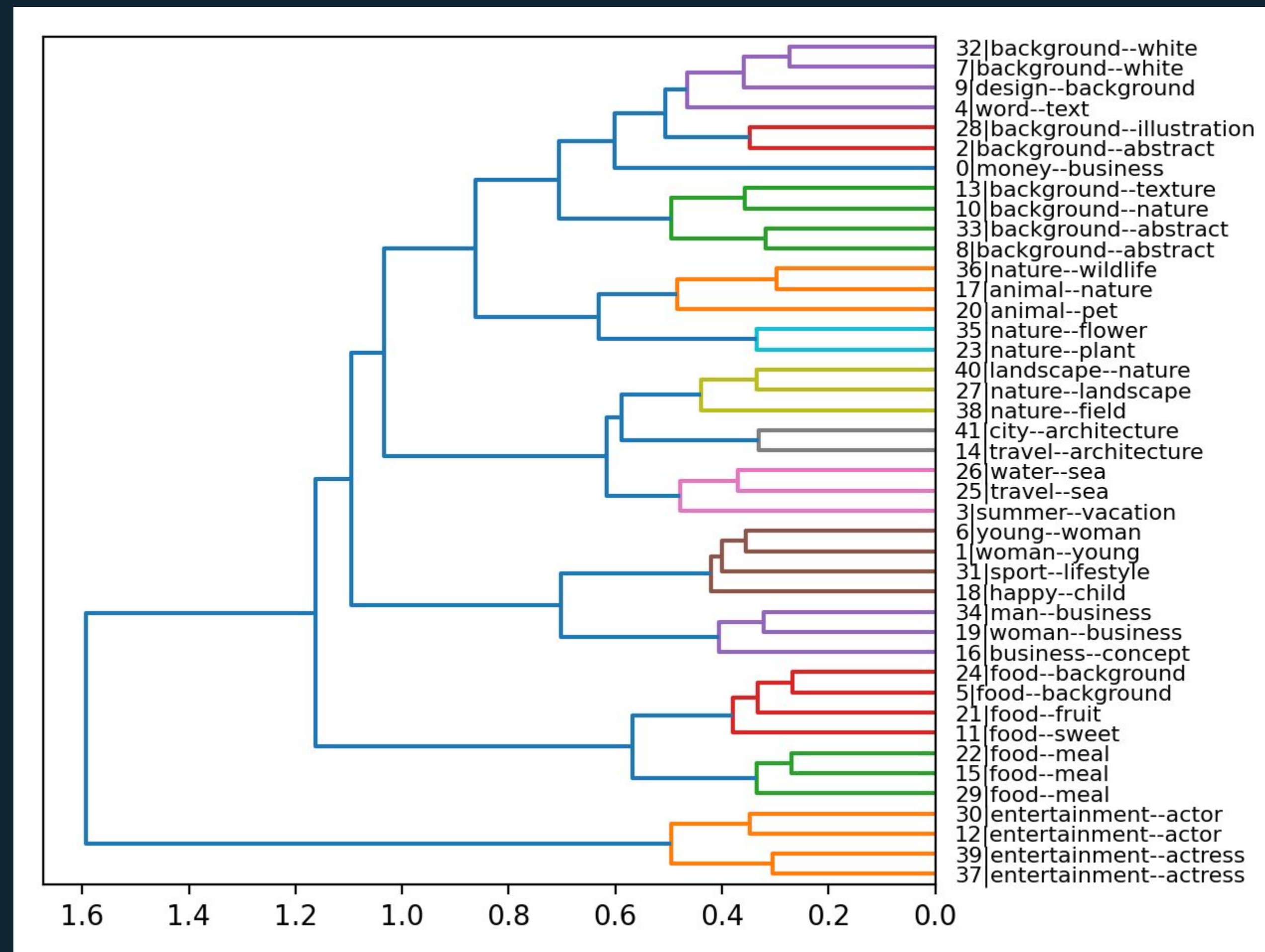


# 4) Cluster-ization





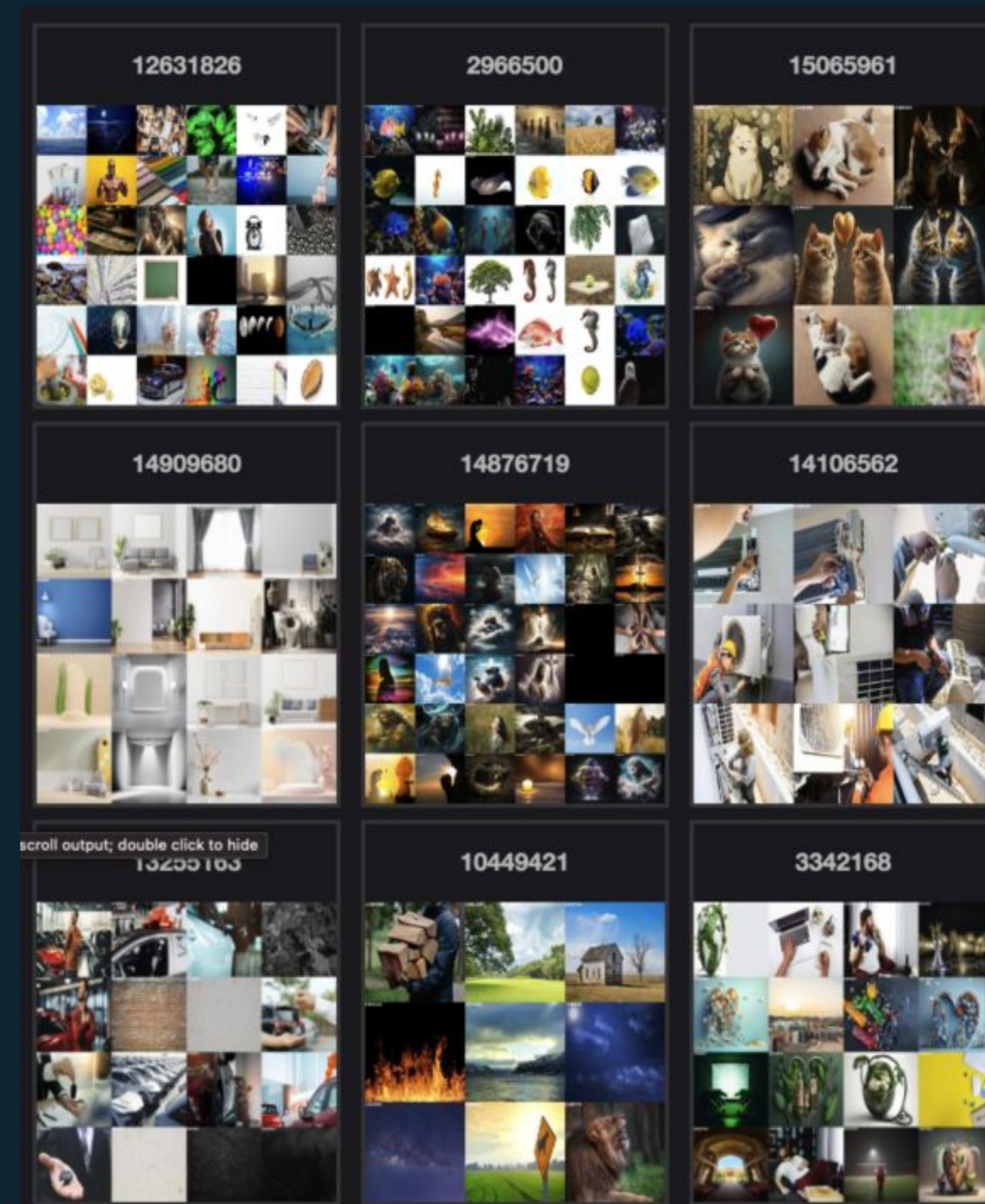
## 4) Cluster-ization





# 4b) Recs2

## User Downloads



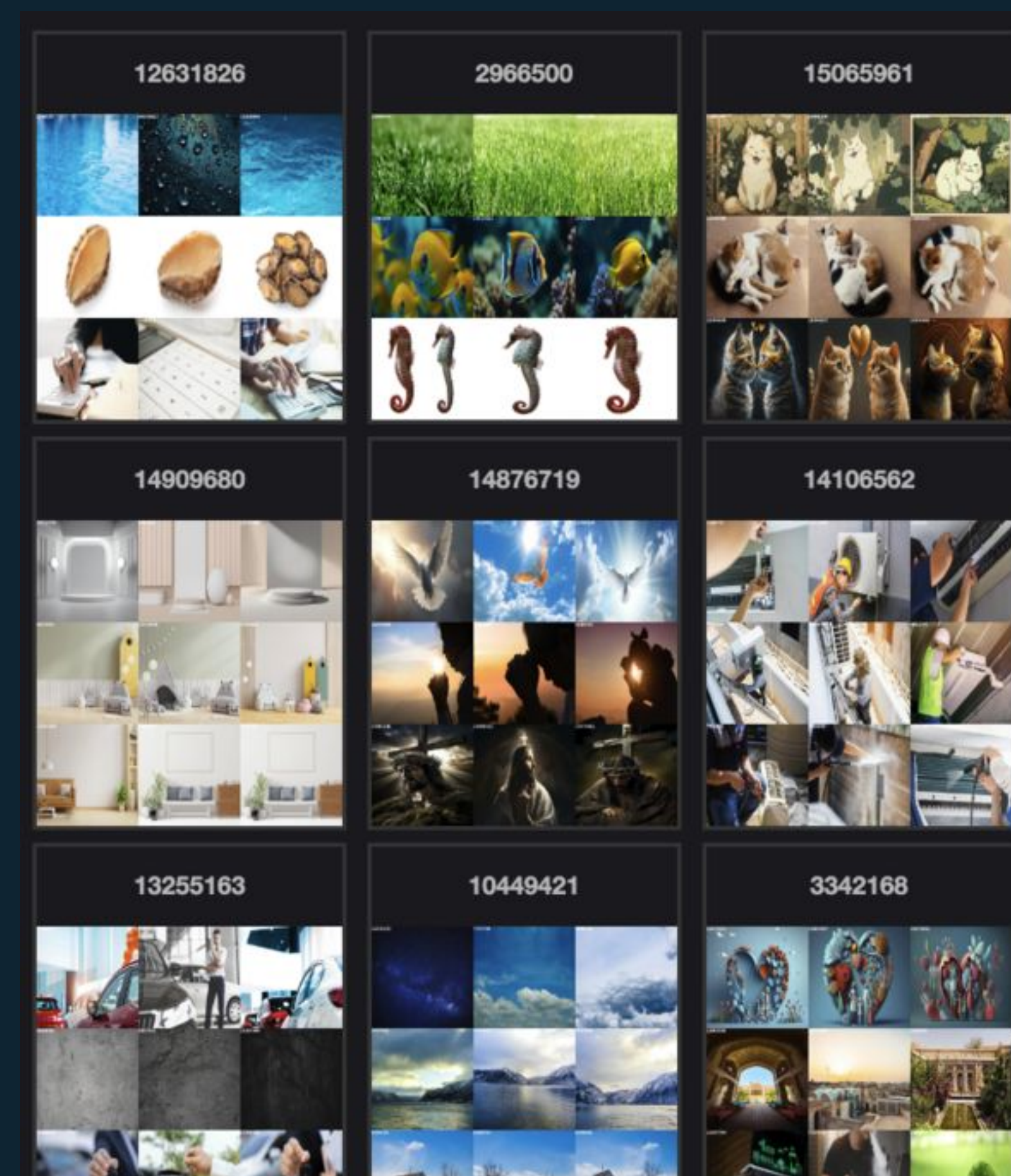
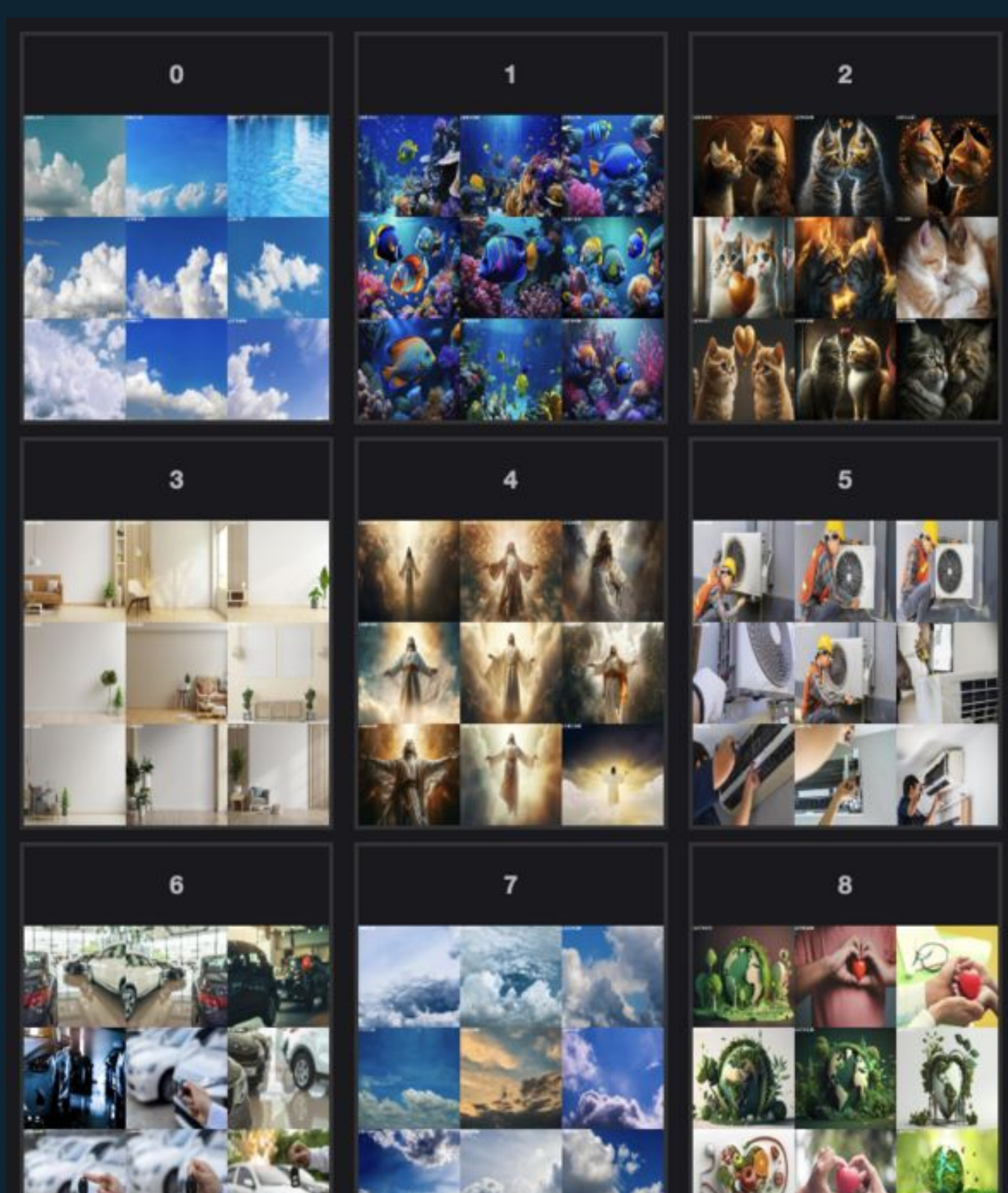


# 4b) Recs2

k=1

Download centroids

k=3



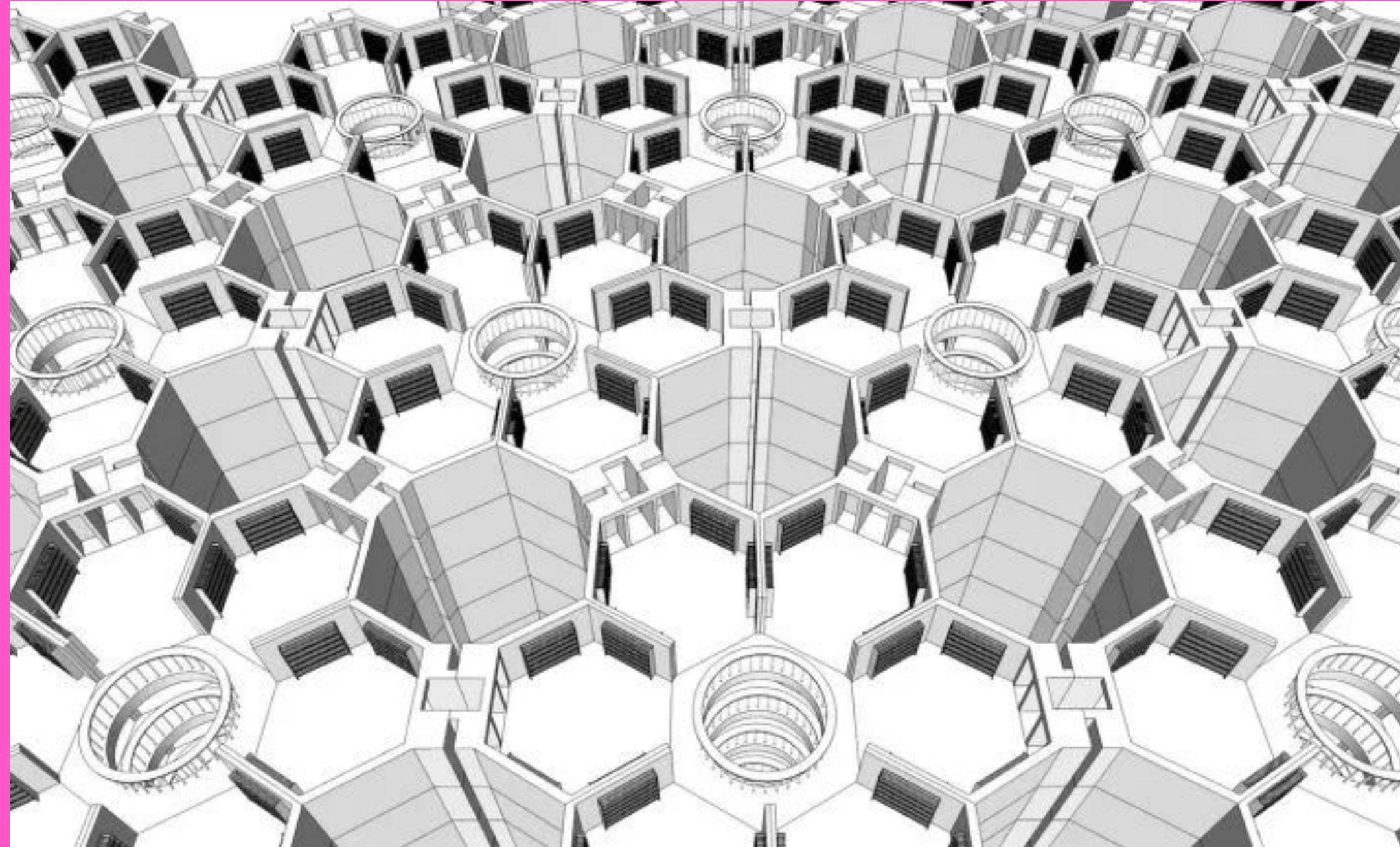


# Talk Timeline



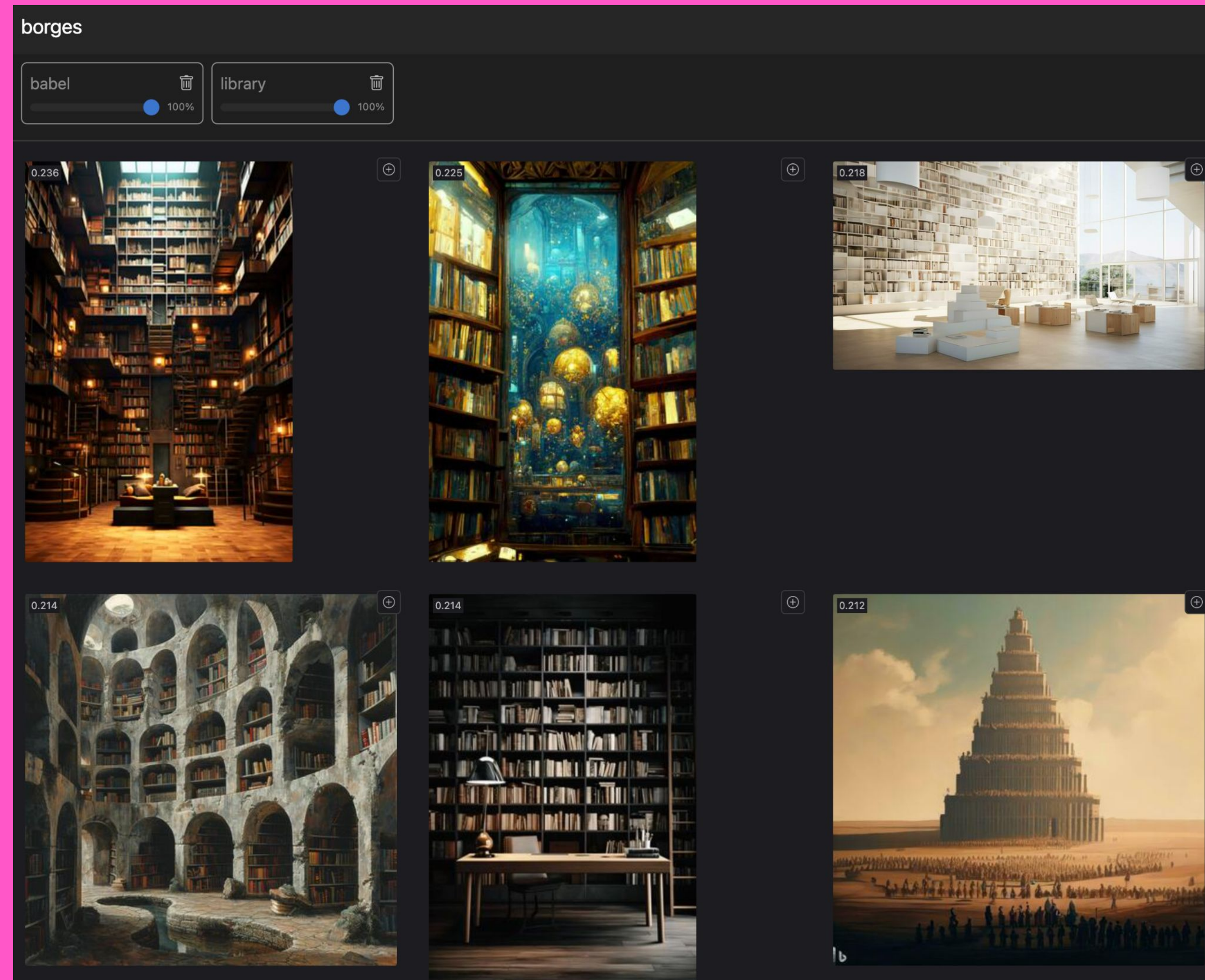


## 5) Jorge Luis Borges





# 5) BORGES





# Talk Timeline

February 2023



Borges  
June 2024

April 2025





# Talk Timeline

