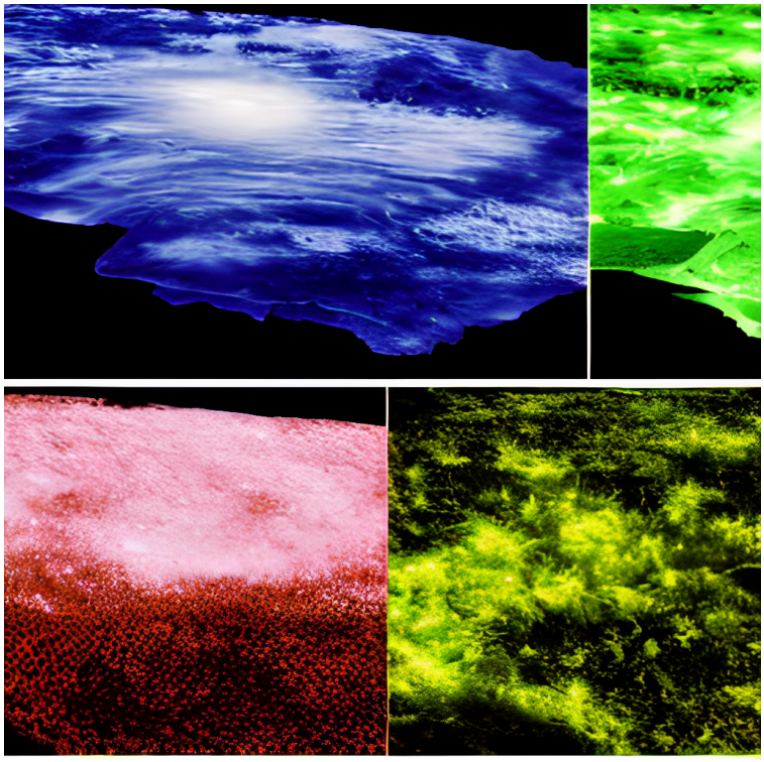
ARTIFICIAL CLIMATE IMAGES

Questioning ***Stable Diffusion's*** interpretations of ***climate change***.



DEN–
SITY
GN+

POLITECNICO
MILANO 1863

SCHOOL OF DESIGN

FINAL SYNTHESIS DESIGN STUDIO
LM in Communication Design
Sez. C3 – 2022/2023

GROUP 06

ARTIFICIAL CLIMATE IMAGES:
Questioning Stable Diffusion’s interpretations of climate change.

AUTHORS
Anna Cattaneo
Lara Macrini
Ce Zheng
Leonardo Puca
Nicole Moreschi
Silvia Sghirinzetti
Yiyuan Hu

FACULTY
Michele Mauri
Ángeles Briones
Gabriele Colombo
Simone Vantini
Salvatore Zingale

TEACHING ASSISTANTS
Elena Aversa
Andrea Benedetti
Tommaso Elli
Beatrice Gobbo
Arianna Bellantuono

DEN–SITY GN+

POLITECNICO MILANO 1863
SCHOOL OF DESIGN

FINAL SYNTHESIS DESIGN STUDIO
LM in Communication Design
Sez. C3 – 2022/2023

Why *AI* and *Climate Change?*

Innovations in the digital sphere have made **Artificial Intelligences** new actors of the communication landscape. Text to image Ais are currently used for artistic purposes, but their possible applications are endless.

What if txt-to-img AIs were in charge of producing climate change imagery?

Nowadays climate change imagery is facing *two dilemmas:*

I ***Mainstream images are not effective in conveying fundamental concepts related to climate change.***

Climate Visual Library* is a remarkable example of what types of images work in climate change communication.

The *Climate Visuals Principles* are used in this report for evaluating the images.

*source
climatevisuals.org

[7 CORE PRINCIPLES]

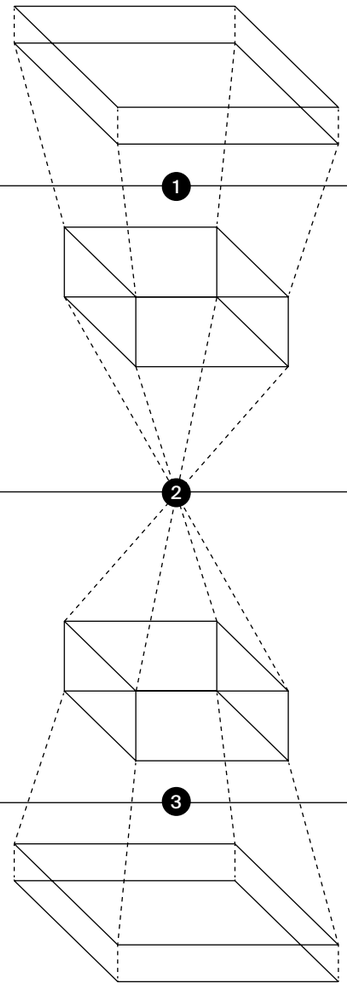
01 Show real people
02 Show climate change causes at scale
03 Understand your audience
04 Tell new stories
05 Show emotionally powerful impacts
06 Show local (but serious) impacts
07 Be careful with protest imagery

II ***The rising technology of text-to-image AIs may soon be in charge of the production of climate change visuals.***

[STABLE DIFFUSION]

This report uses **Stable Diffusion**, an open source txt-to-img AI, to question the relationship between human and machine in the context of climate change imagery.

Stable Diffusion works in **three main steps**; this report questions each of them:



[TRAINING SET]

Humans give an enormous amount of images to the AI, this is everything the AI learns about climate change.

[QUESTION 1]

Focus on the training set in order to understand what information is given to the AI.

[LATENT SPACE]

The machine processes the knowledge and is able to produce images, but such process is a black box, unknown by humans.

[QUESTION 2]

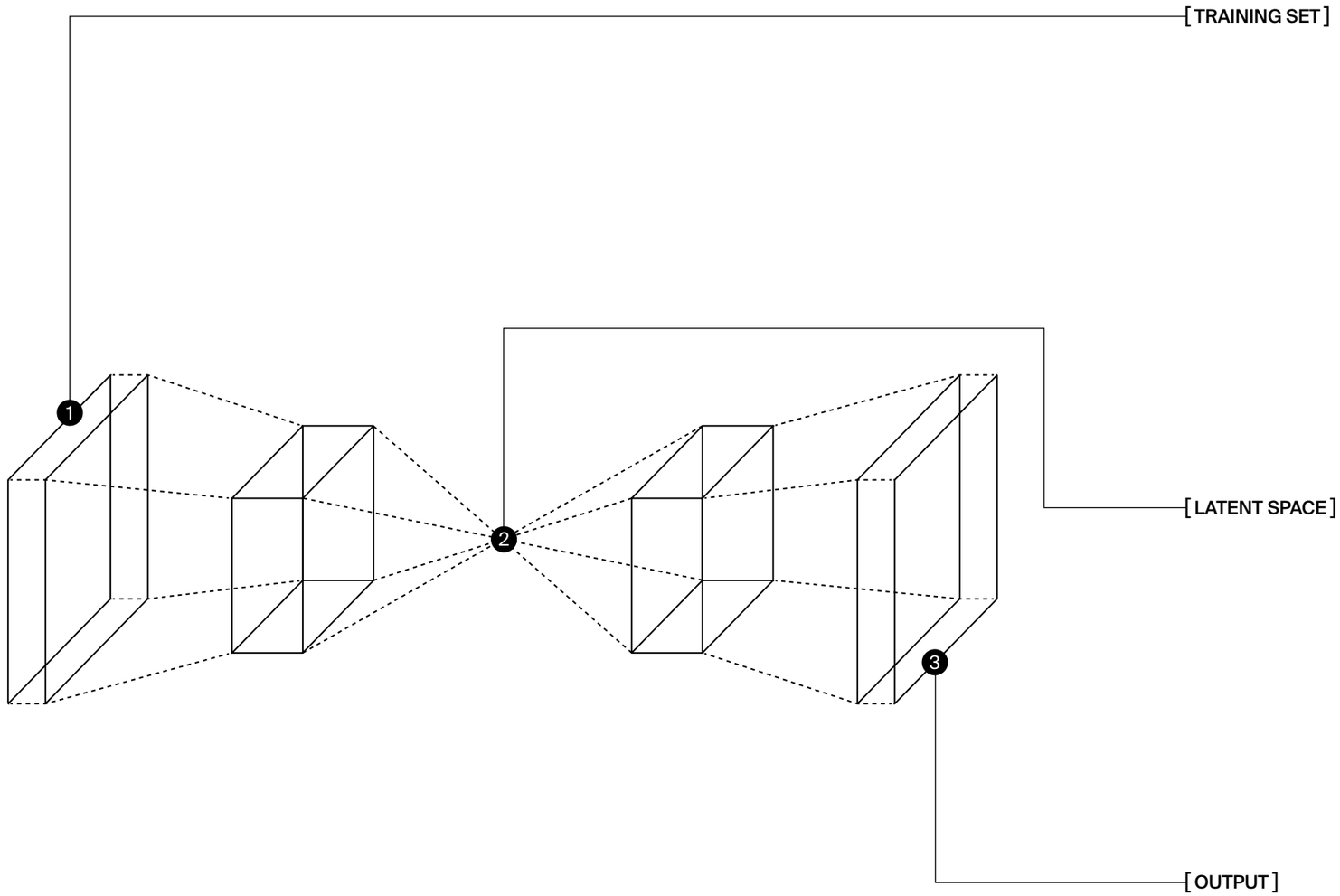
Focus on latent space in order to grasp how the generating process occurs.

[OUTPUT]

Humans and human related issues to climate change are depicted by the AI.

[QUESTION 3]

Focus on output images to evaluate what the AI is able to tell about climate change and its actors.



How are climate change actors and actions depicted in Stable Diffusion's training set?

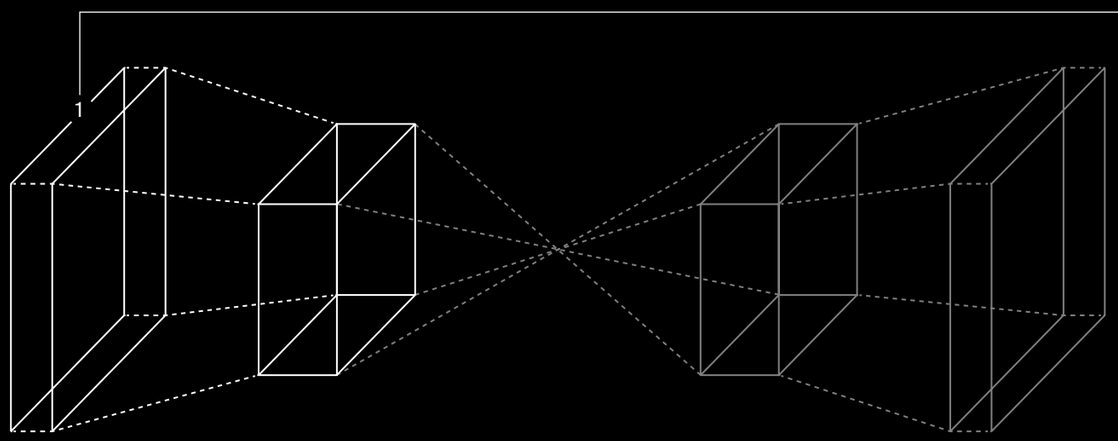
RESEARCH PROTOCOL 10
INTRODUCTION TO TRAINING SET 11
SELECTED IMAGES 12
IMAGE CLUSTERING 14
ETHNICITY - ACTION 16
KNOWN PEOPLE - ACTION 18
EFFECTS - CAUSES - SOLUTIONS 20
CLICHÈ IMAGES 22

What can latent space tell us about Stable Diffusion's interpretations of climate change?

RESEARCH PROTOCOL 26
INTRODUCTION TO LATENT SPACE 27
TOPIC AREAS 28
TOPICS INTERPOLATION 36
MOST RELEVANT IMAGES 42

How does Stable Diffusion represent people and their actions related to climate change in generated images?

RESEARCH PROTOCOL 50
INTRODUCTION TO OUTPUT 51
GENERATED IMAGES OVERVIEW 52
PEOPLE IN GENERAL PROMPTS 56
IMPACTS AND SOLUTIONS ACTORS 58
SOLUTIONS ACTIONS 60
CONTEXTUALIZATION 62



How are climate change actors and actions depicted in Stable Diffusion's training set?

How to read it:

SECTIONS

STEPS

[TOOLS]

1 DATA COLLECTION

[LAION AESTHETIC] — Select the training set

Select the prompt — Climate change

Define the sorting criterias — First 50 img by similiarity; First 50 img by aesthetic

[CHROME IMAGEYE] — Download images

2 DATA ANALYSIS

Define the clusters based on: — Climate visual principles; Observation of the dataset

[EXCEL] — Sort the images — Medium; Actors; Actions

3 DATA VISUALIZATION

[RAWGRAPHS] — Overall visualization

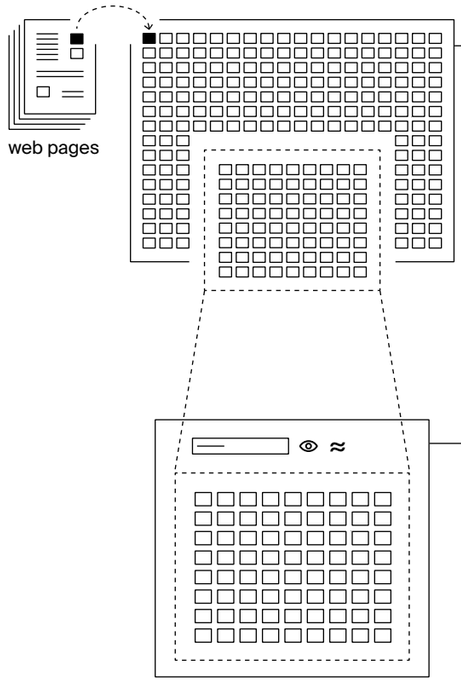
Zoom in relationships actors/actions → Finding 1

Zoom in proportions known people/actions → Finding 2

Zoom in causes/impacts/solutions → Finding 3

Zoom in clichè images → Finding 4

Stable Diffusion is trained on *Laion-5b* dataset.



[LAION-5B] The largest available **dataset** of **image-caption** pairs (5.85 billion). The images in **LAION-5B** dataset are scraped **from the web** (Common Craw).

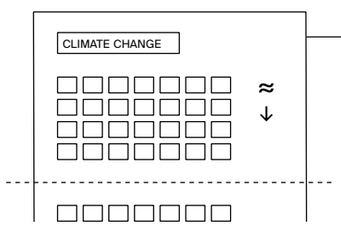
Each image is paired with a **caption**, directly taken from its alt attribute in the html.

[LAION-AESTETHIC] **LAION-Aesthetics** is a 600 million image subset of LAION-5B, composed by images with a hight aesthetics score.

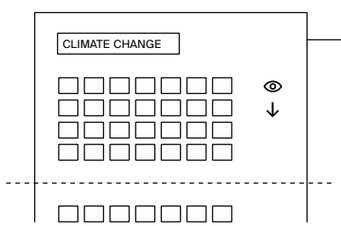
A data browser allows to explore it by **queries** (CLIP model) and **sorting criterias**.

Images collection from *Laion-Aesthetic*

Usage of the filters "**similarity"** and "**aesthetic score"** to sort the images.



Collection of the top **50 images** that answer the query "**climate change**", sorted by **similarity value** (degree of similarity between the image and caption, as measured using the CLIP model).



Collection of the top **50 images** that answer the query: "**climate change**", sorted by **aestethic value** (stimated score that a human would assign to the aesthetics of the image.).

≈ 50 img + 50 img = **100 img**

SELECTED IMAGES

IMAGE COLLECTION 1

Dataset: Laion Aesthetic
Sample: 50 images
Sorted by: Similarity Value

[TOP 50 BY SIMILARITY]

IMAGE COLLECTION 2

Dataset: Laion Aesthetic
Sample: 50 images
Sorted by: Aesthetic Value

[TOP 50 BY AESTETHIC]

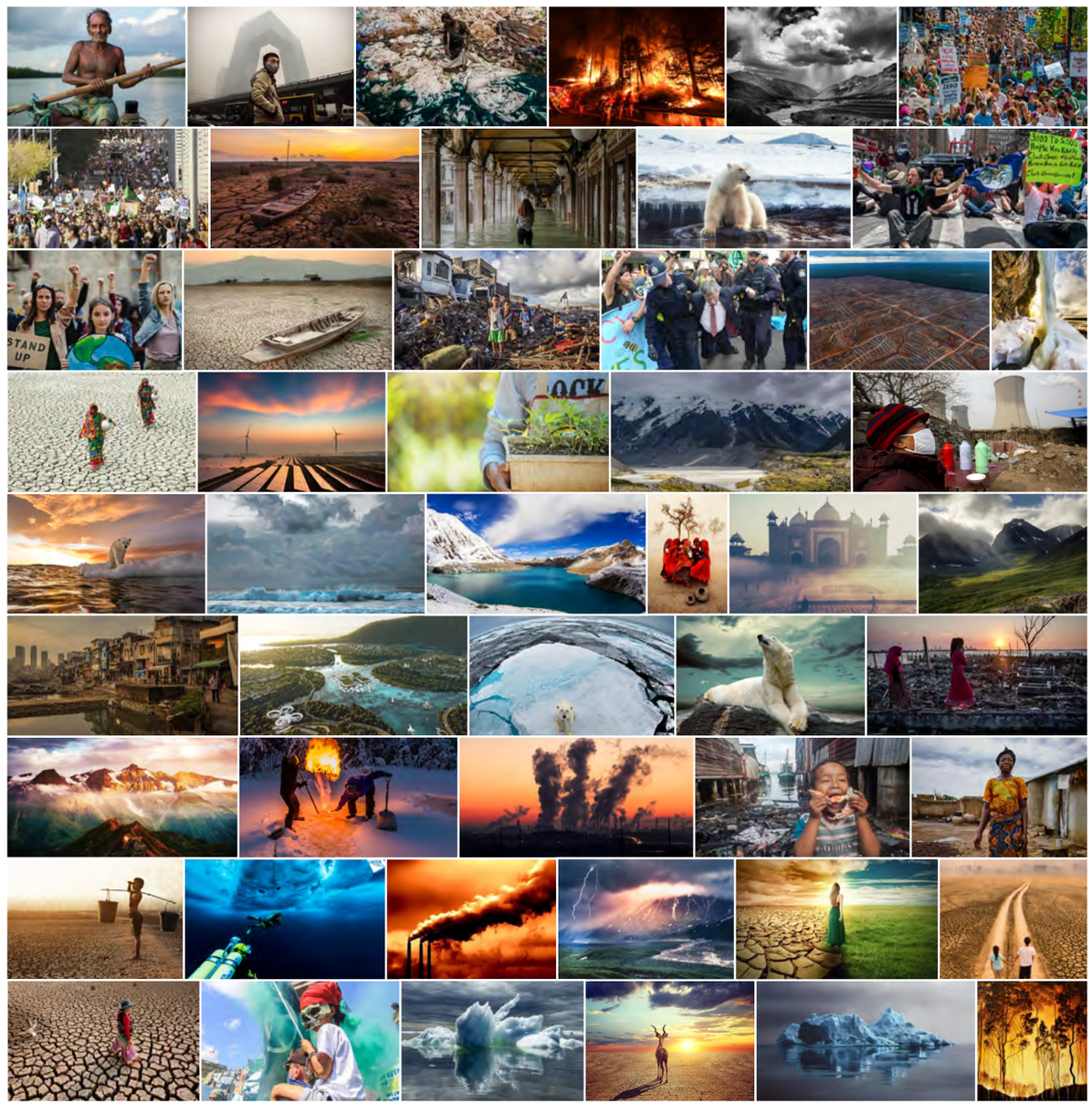
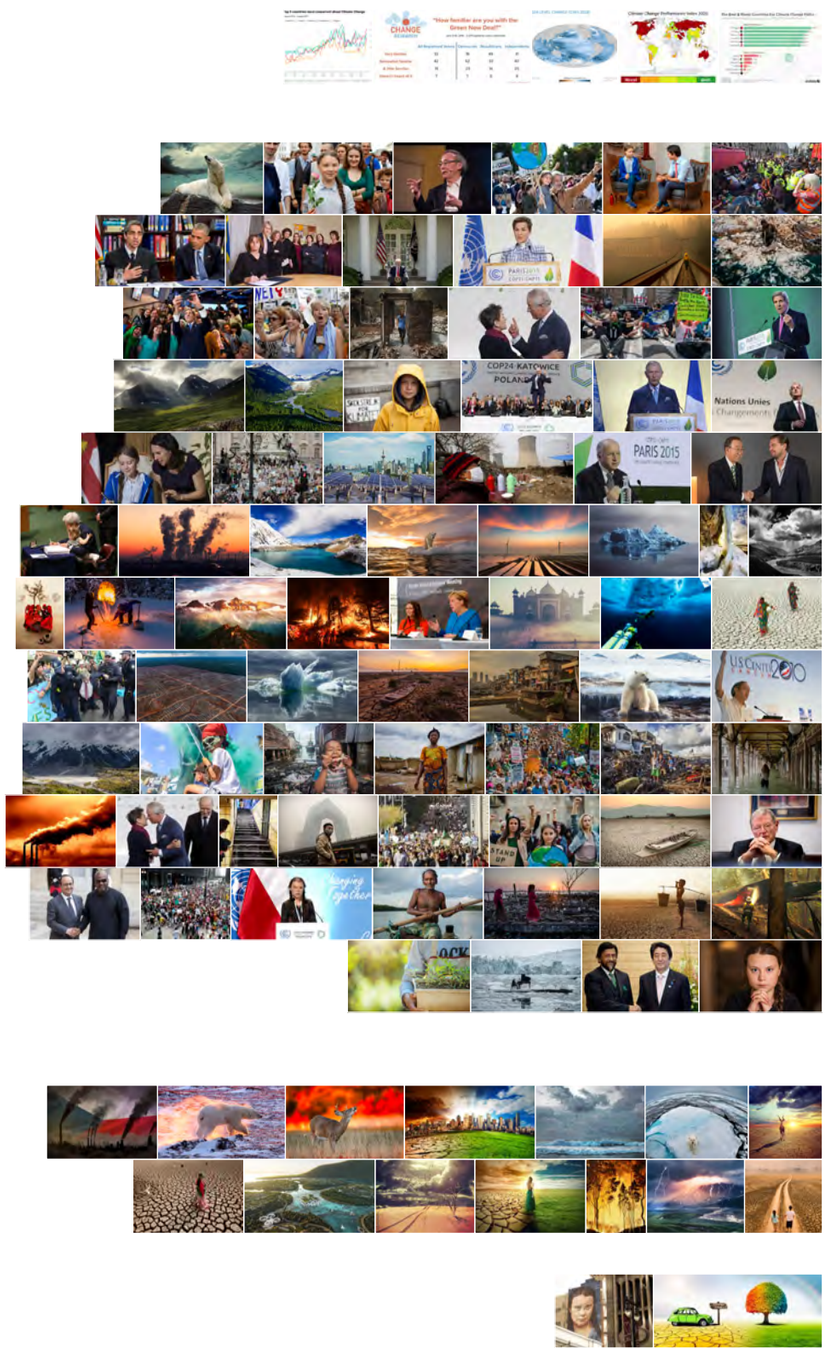
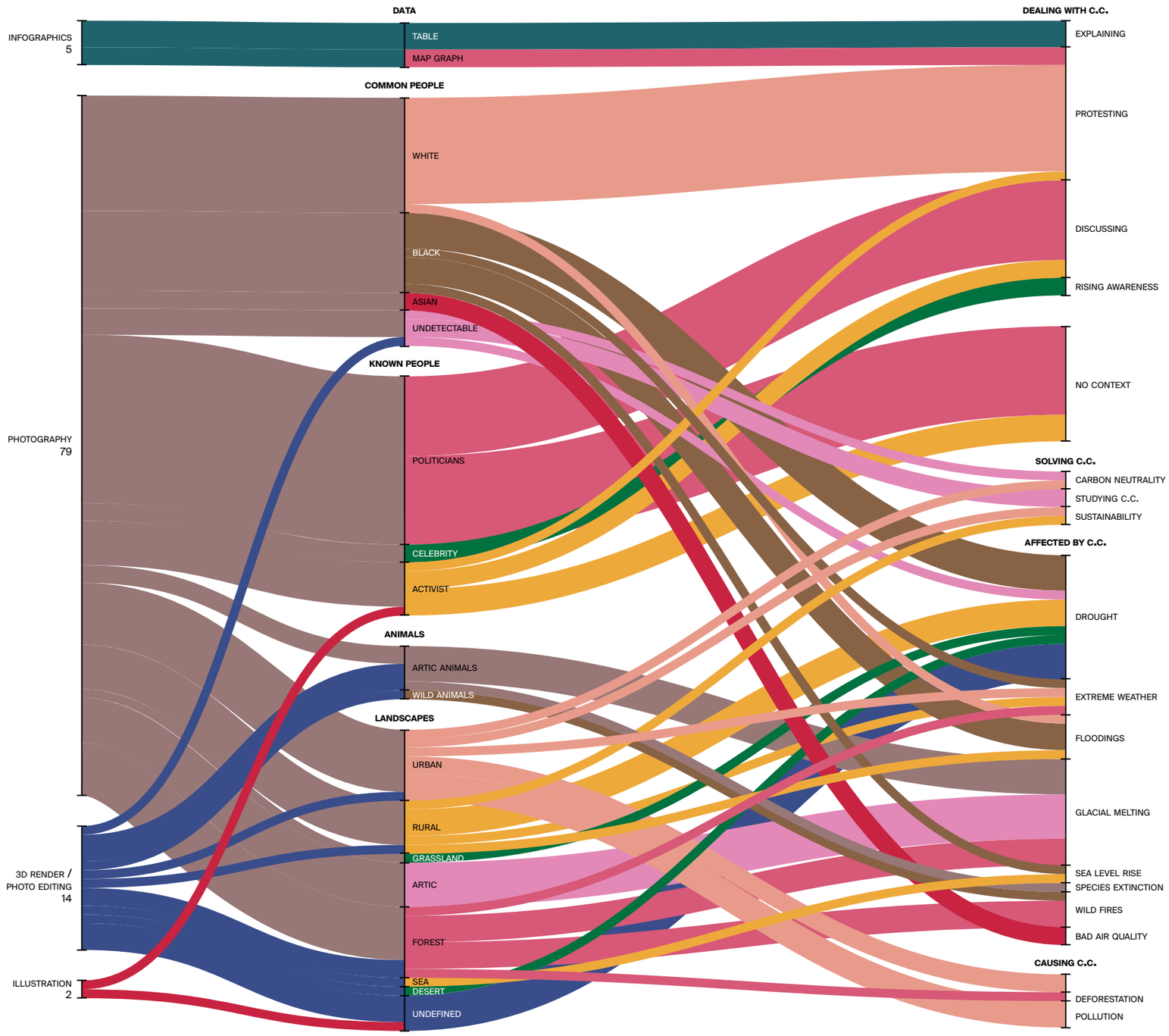


IMAGE CLUSTERING

OVERALL VISUALIZATION

The 100 images are clustered by medium, actors and actions. The aim is to analyze their relations.

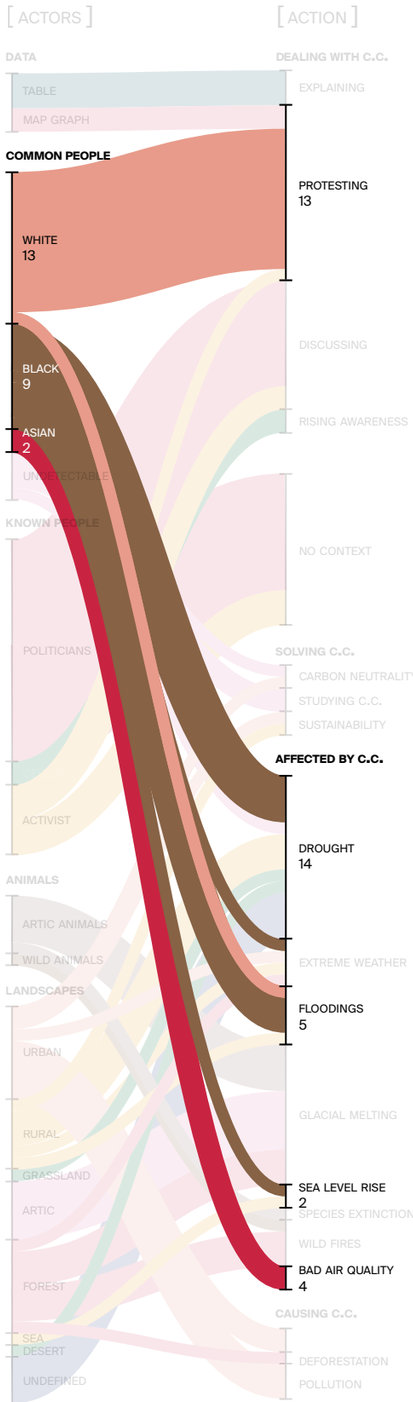




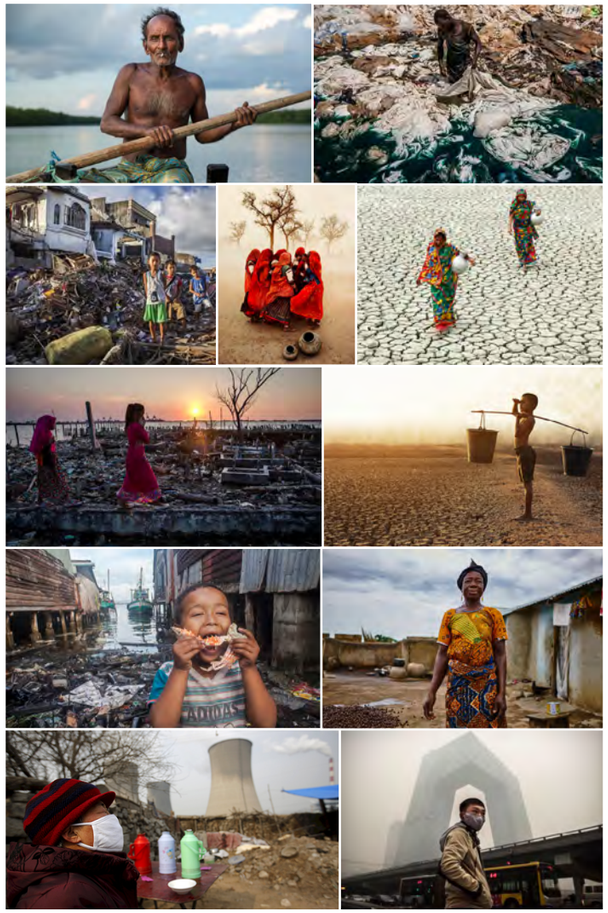
ETHNICITY – ACTION

ETHNICITY BIAS

Analisys of the relationships between the ethnicity and the action of the actors involved.



[BLACK | ASIAN PEOPLE: AFFECTED BY CLIMATE CHANGE] [tot. 11]



[WHITE PEOPLE: PROTESTING FOR CLIMATE CHANGE] [tot. 12]

[FINDING] ❶

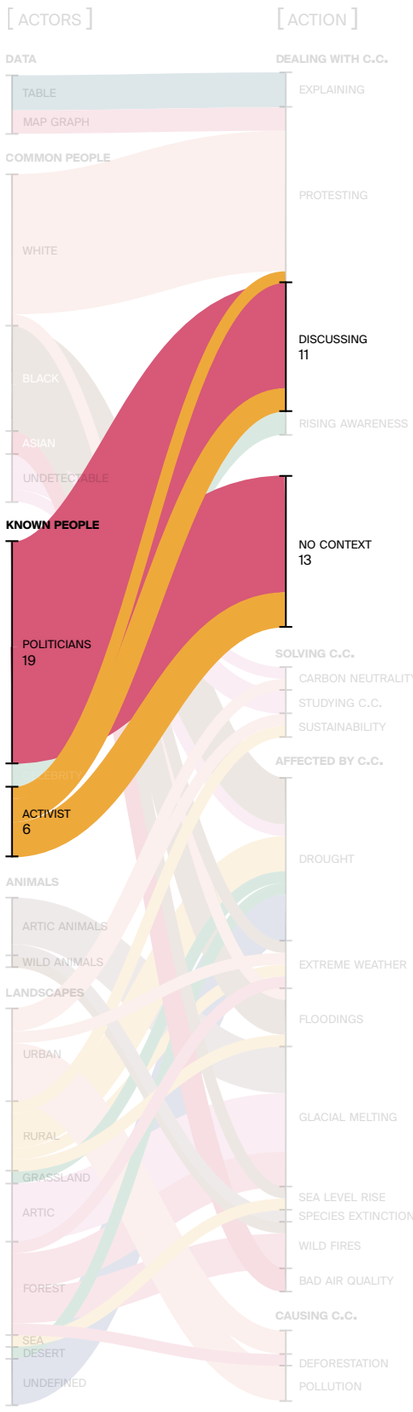
Almost all ***white people*** are ***protesting*** for climate change, while all ***other ethnicities*** are being ***effected*** by it.

Among the 13 images that present white people, 12 of them depict protests for climate. However, in all the 11 images that contain asian or black people, they are suffering the impacts of climate change (drought, floodings, sea level rise and bad air quality).

KNOWN PEOPLE – ACTION

POLITICIANS AND ACTIVISTS

Analisys of the amount of public figures and their relations with climate change.



[POLITICIANS I ACTIVISTS: DISCUSSING] [tot. 10]

[POLITICIANS I ACTIVISTS: UNDEFINED CONTEXT] [tot. 15]

[FINDING] ❷

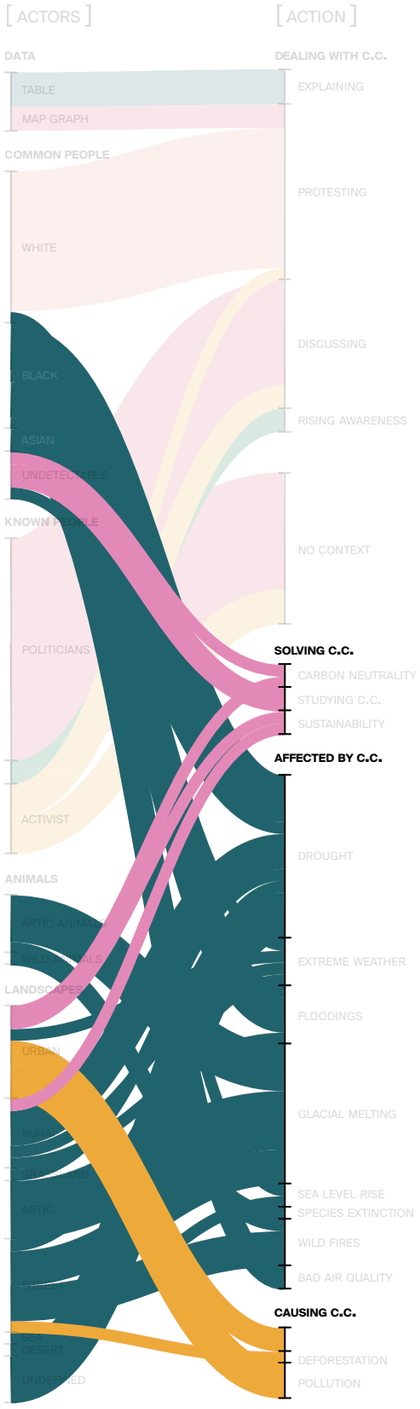
Known people – ***politicians*** and ***activists*** – are very ***popular*** in the dataset, even if the climate change context is not easily detectable.

Images of politicians and climate change activists are widely present in the dataset (25 images). While some of them are depicted discussing about climate change (detectable from the texts), most of them completely lack the context.
Staged images of polititians are seen as less 'authentic', even manipulative, so they are not a good climate change visual (based on the 1st principle of climate visuals).

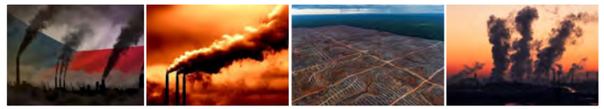
EFFECTS - CAUSES - SOLUTIONS

LACK OF CAUSES AND SOLUTIONS

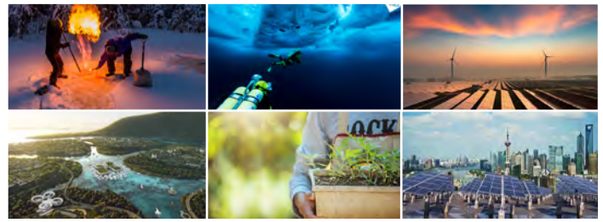
Analisys of amount of climate change effects imags compared with causes and solutions.



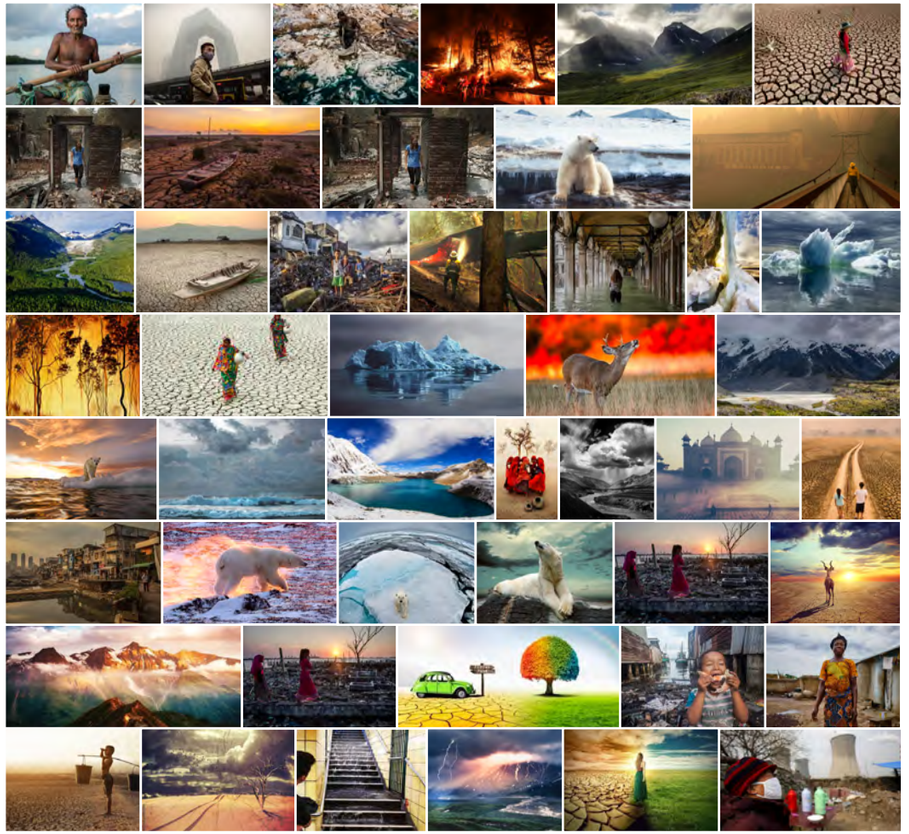
[CAUSING CLIMATE CHANGE] [tot. 4]



[PROVIDING SOLUTIONS TO CLIMATE CHANGE] [tot. 6]



[EFFECTED BY CLIMATE CHANGE] [tot. 47]



[FINDING] ❸

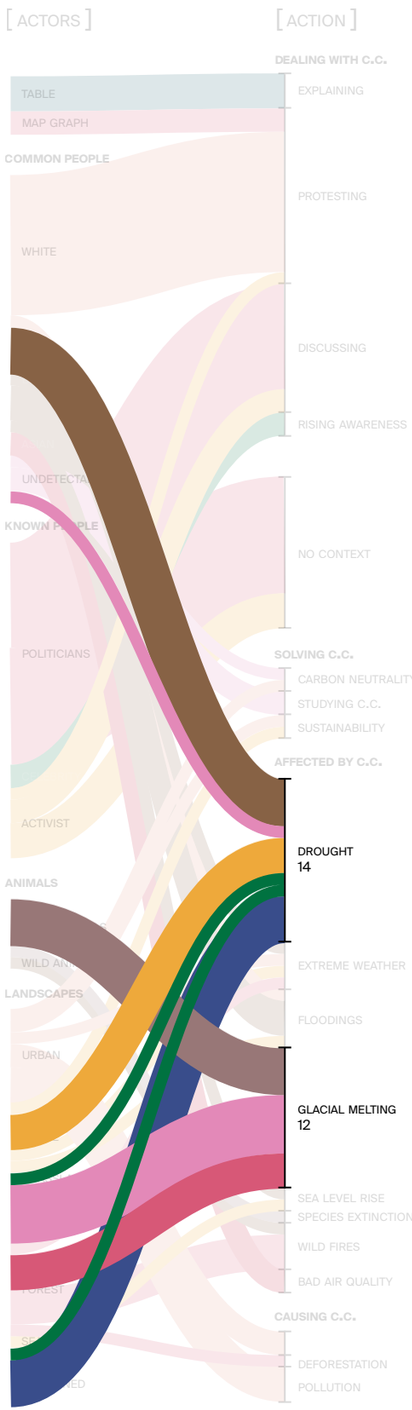
The majority of the images show the ***effects*** of climate change. Just a few of them depict ***causes*** or ***solutions***.

Almost half of the images collected (47) represent the impacts of climate change. Only few of the actually provides visualizations of solutions (6) or causes (4).
Images of climate impacts can prompt a desire to respond, but because they are emotionally powerful, they can also be overwhelming (4th principle of climate visual).

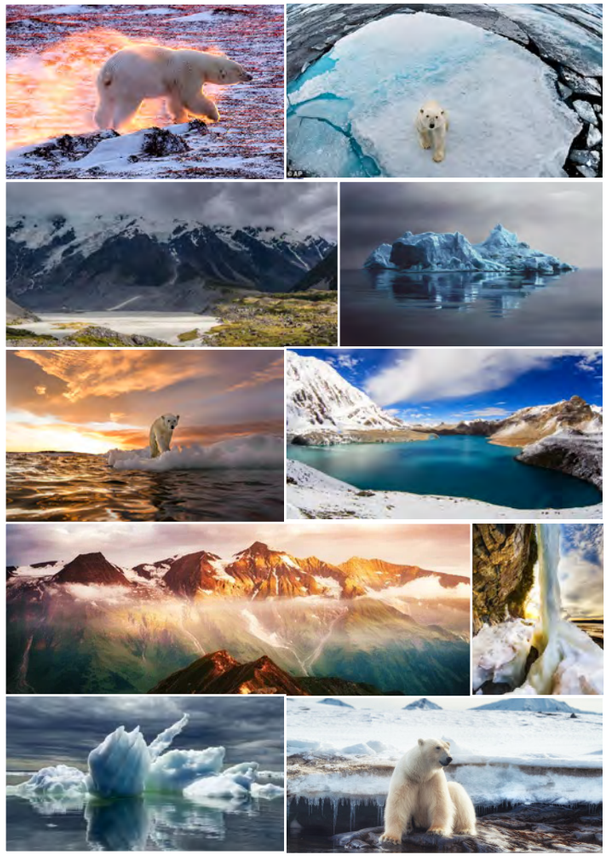
CLICHÈ IMAGES

DROUGHT AND ICE

Analisys of the quantity of familiar, overused images about climate change.



[GLACIAL MELTING] [tot. 10]

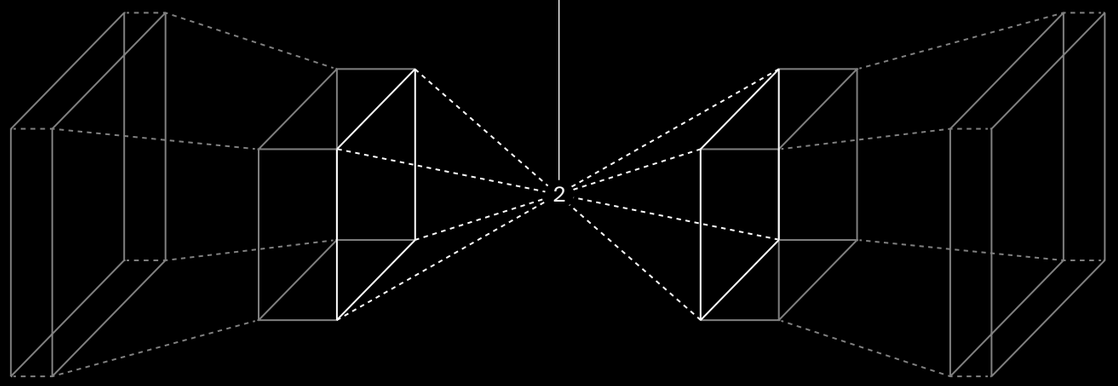


[DROUGHT] [tot. 12]

[FINDING] 4

A great deal of imagery is ***cliché***: the most common are ***melting glaciers*** and ***drought***, often exaggerated in an unrealistic style.

The dataset contains lots of clichè images, mainly regarding drought (12) and melting ice (10). "Classic" images may be an effective way of communicating the topic climate change, but less familiar (and more thought-provoking) images can help tell new stories and remake the visual representation o climate change in the public mind (2nd principle of climate visuals).



[LATENT SPACE]

What can latent space tell us about Stable Diffusion's interpretations of climate change?

How to read it:

SECTIONS

STEPS

[TOOLS]

1 DATA COLLECTION

Explore different CFG and CLIP values

[AUTOMATIC1111 / STABLE DIFFUSION WEB UI] — Generate different X/Y PLOTS

Collect the images — [56 images per plot] [18 plots generated]

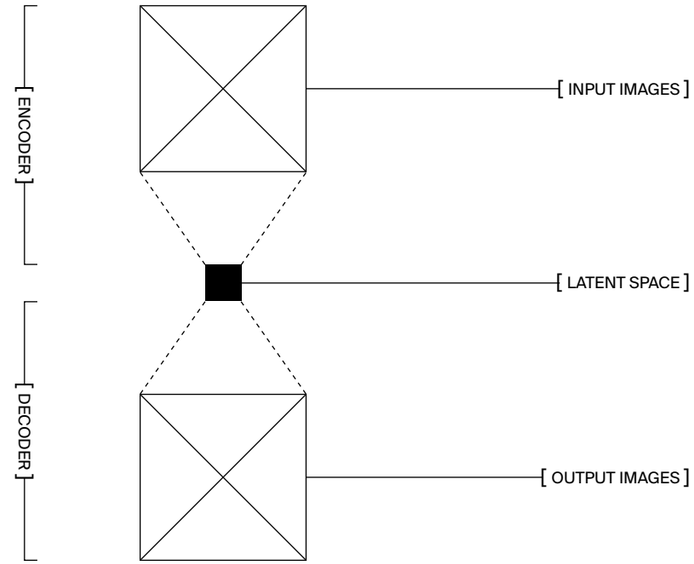
2 DATA ANALYSIS

Qualitative analysis based on direct observation

3 DATA VISUALIZATION

- TOPIC AREAS → Finding 1
- TOPICS INTERPOLATION → Finding 2
- RELEVANT IMAGES → Finding 3 - 4

latent space is a representation of compressed data.



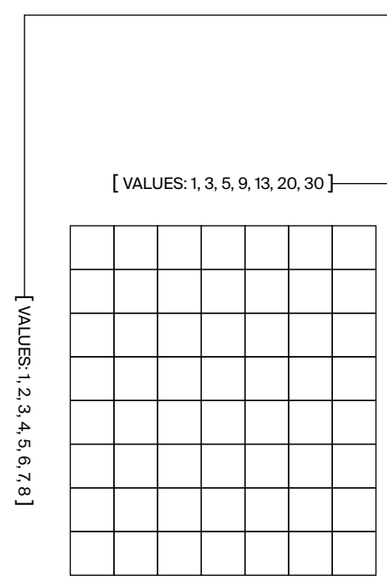
Training a deep learning model with images means teaching the model to **identify similarities** between them.

Latent space contains a representation of the images. It stores all the **relevant features** learned from the training set.

Images generated from the model allow to **find patterns** and understand the relevant features stored in the latent space.

We can ***explore*** the latent space through an ***x/y plot*** of the outputs.

The x/y plot is composed by 56 images generated and sorted by following 2 settings:



Contrastive Language–Image Pre-training: it affects **how images and text are connected**. The values allows to visualize a representative evolution of images.

Classifier Free Guidance: it affects **how closely the model follows the prompt**. The chosen range of values allows to generate significant images and avoid bad artifacts.

[SETTINGS]

Each image within the plot was generated from the prompt: “**climate change**”

Seed: random
Sampling method: DDIM
Steps: 30
Size: 512x512 px

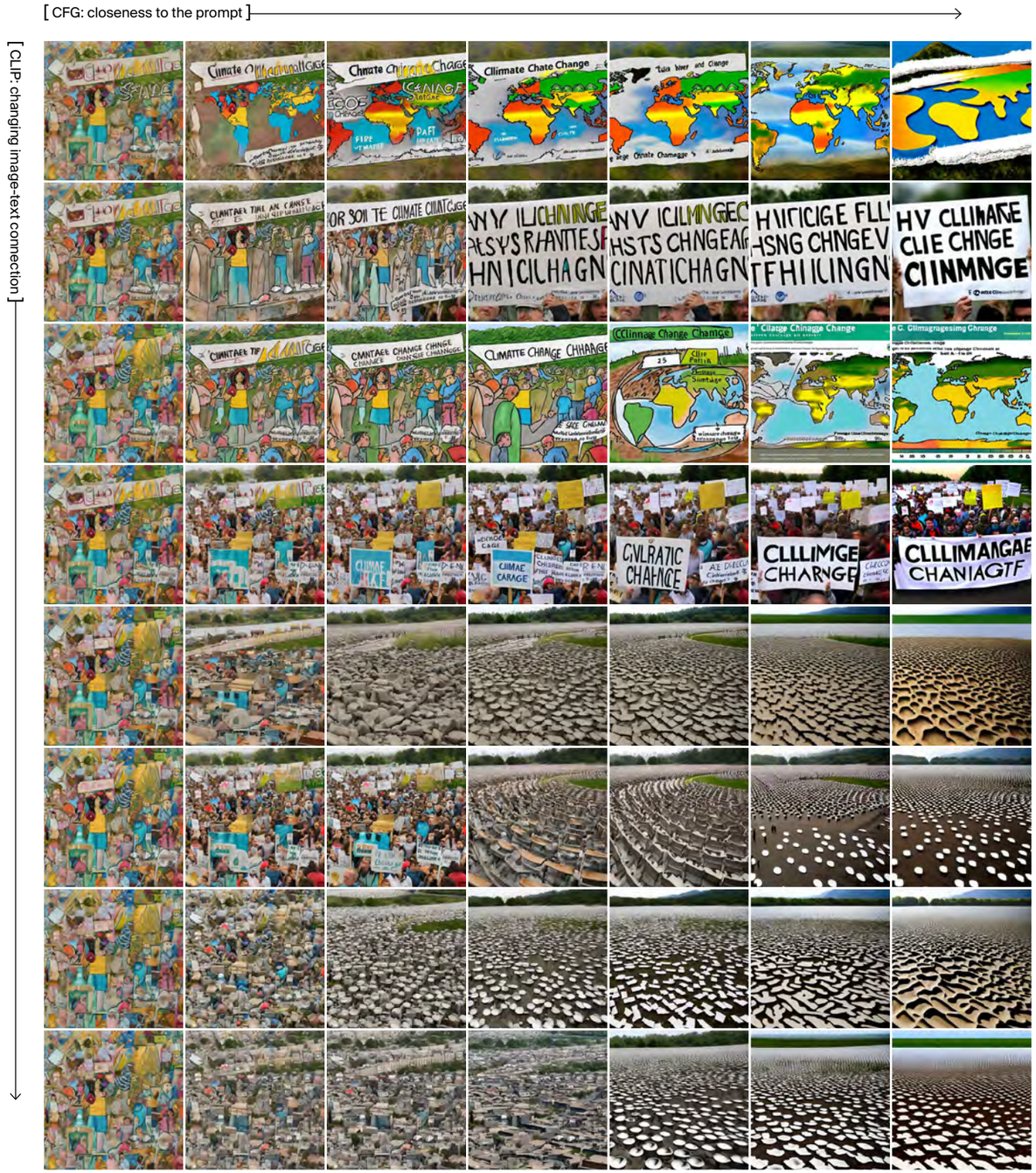
TOPIC AREAS

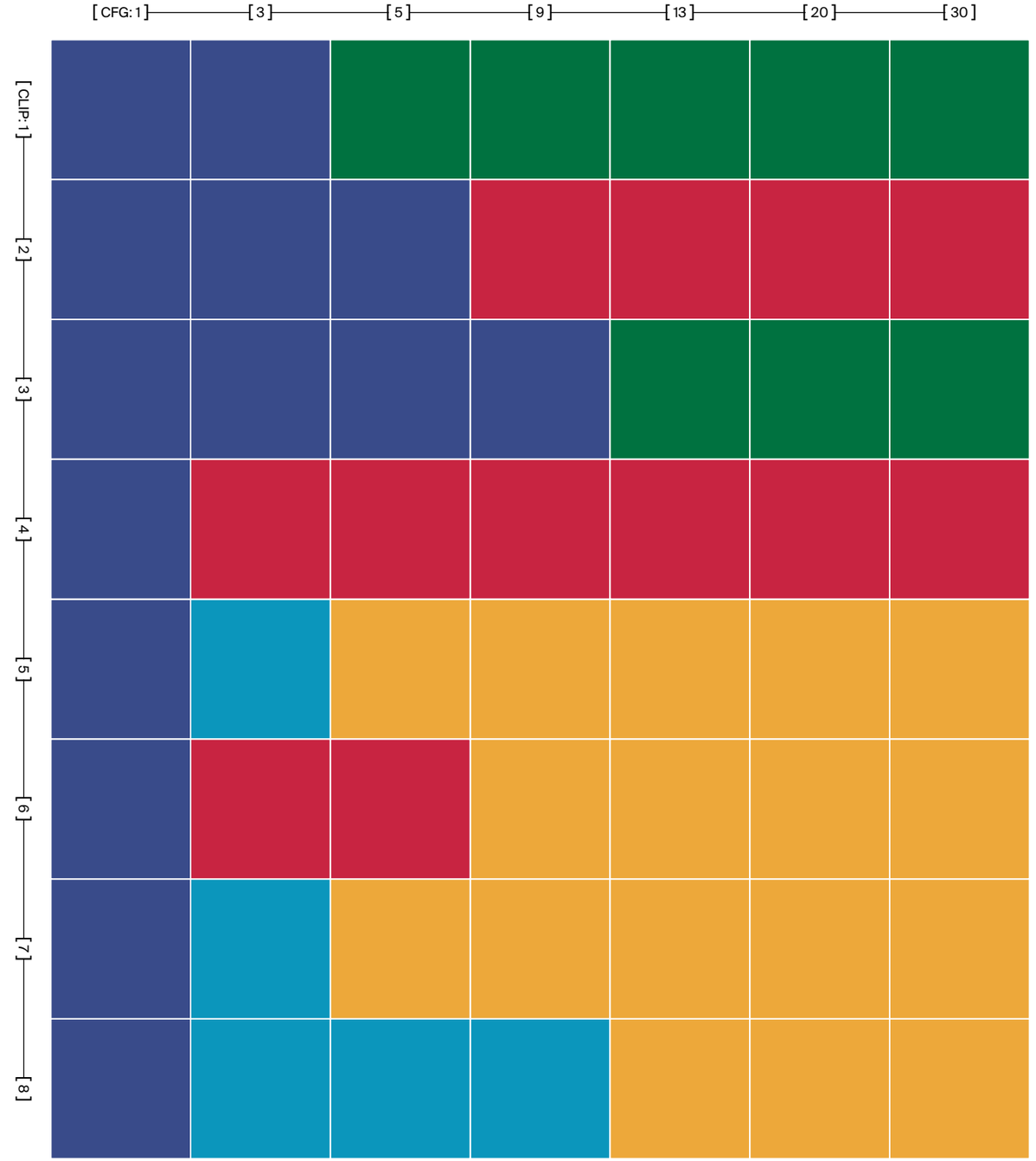
X/Y PLOT ANALYSIS:

Each plot was analyzed in order to identify climate change related topics and how they are related to different CFG and CLIP values.

How to read it:

NOISE
URBAN LANDS
PROTEST
PEOPLE
NATURAL LANDS
INFOGRAPHICS





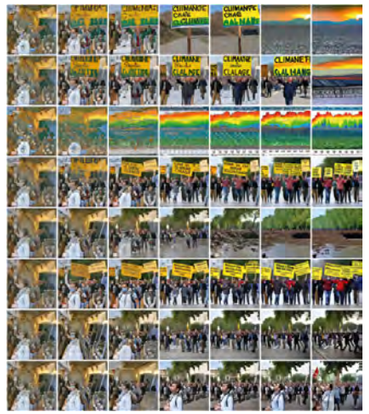
TOPIC AREAS

X/Y PLOT OVERVIEW:

18 different plots were generated from the same prompt: “climate change”

[1]

[2]

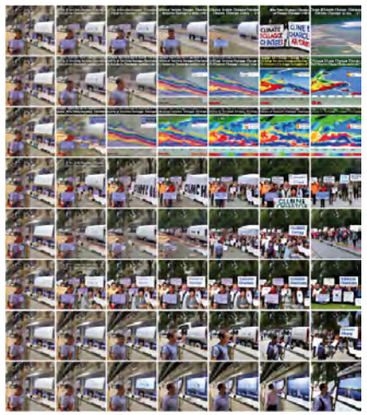


[3]

[4]

[5]

[6]

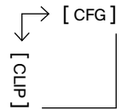


[7]

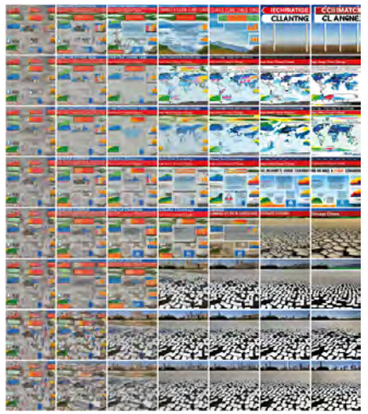
[8]

[9]

How to read it:



[10]



[11]

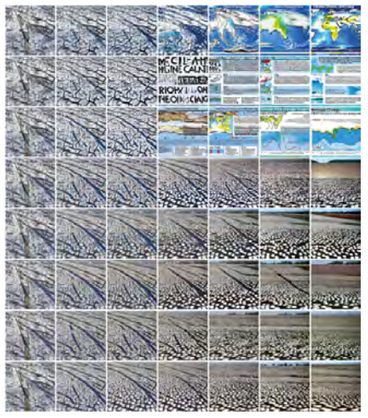
[12]

[13]

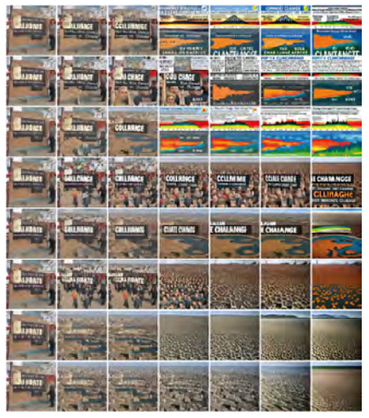
[14]

[15]

[16]



[17]



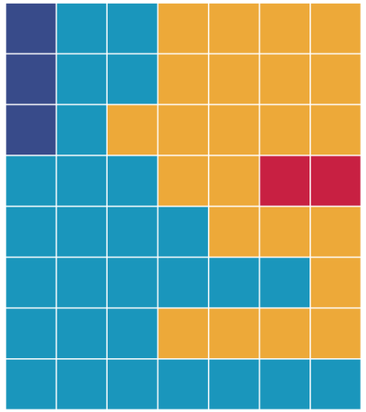
[18]

TOPIC AREAS

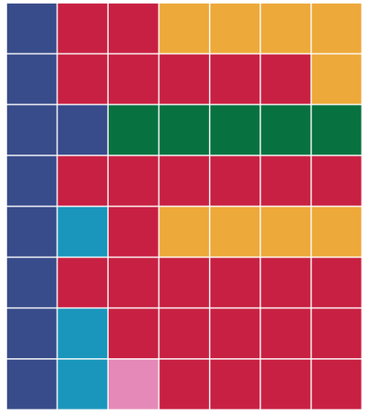
X/Y PLOT OVERVIEW:

The method was then applied to all the plots and 6 main recurring topics were identified within them.

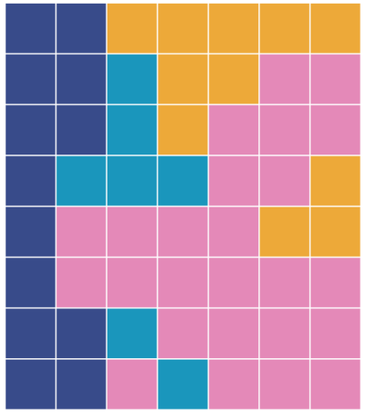
[1]



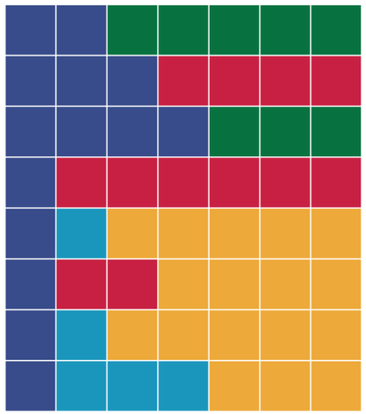
[2]



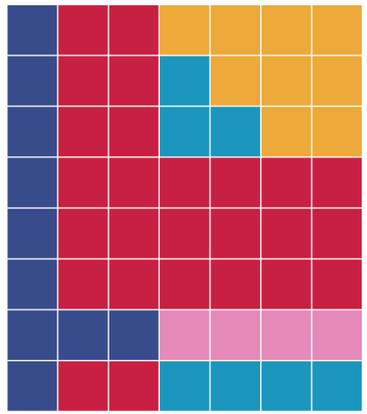
[3]



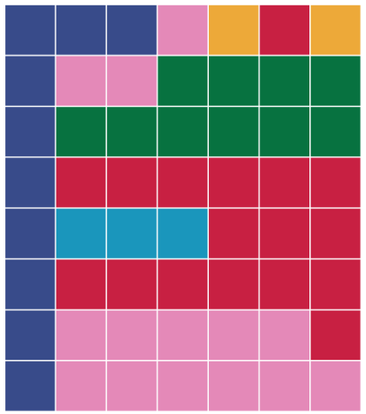
[4]



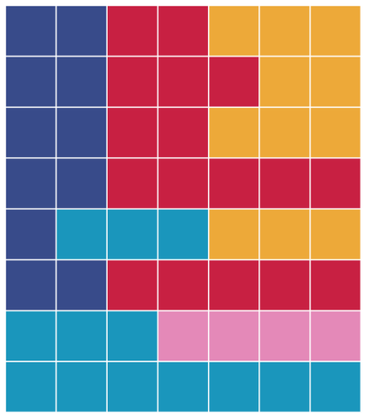
[5]



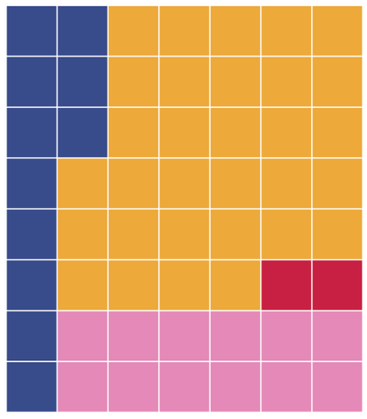
[6]



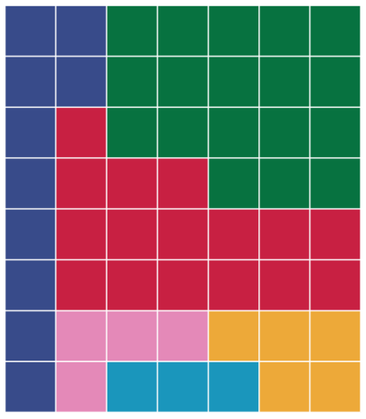
[7]



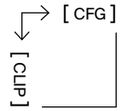
[8]



[9]

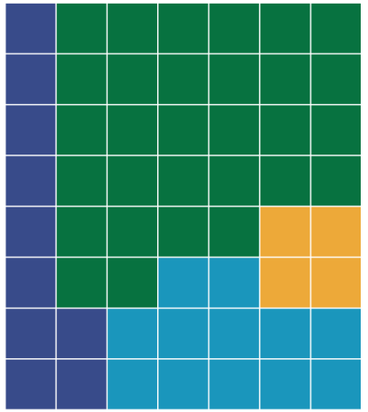


How to read it:

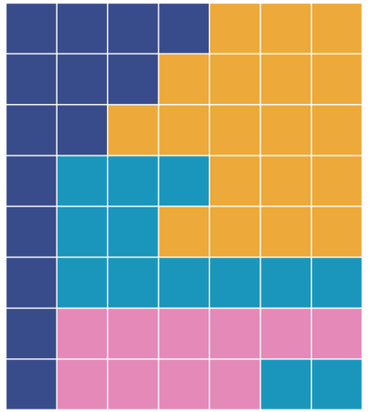


NOISE
URBAN LANDS
PROTEST
PEOPLE
NATURAL LANDS
INFOGRAPHICS

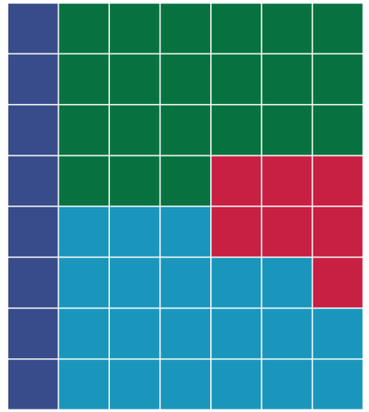
[10]



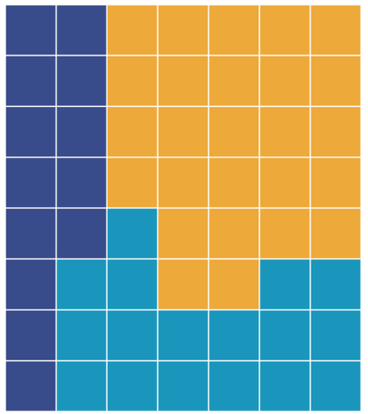
[11]



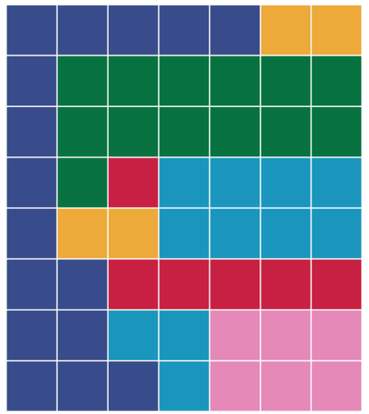
[12]



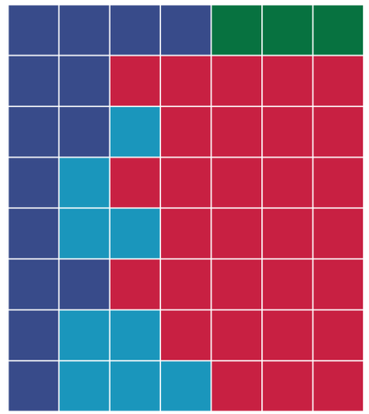
[13]



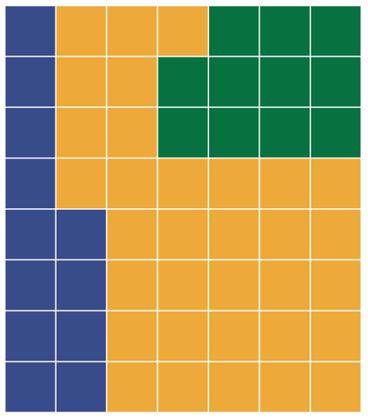
[14]



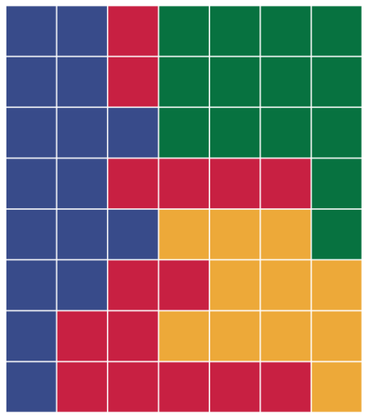
[15]



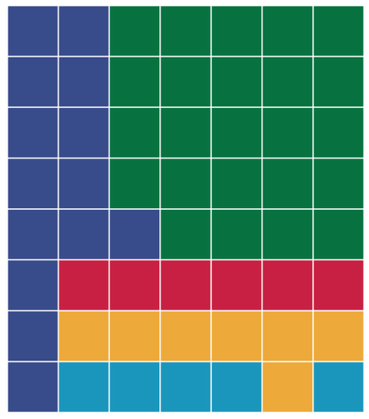
[16]



[17]

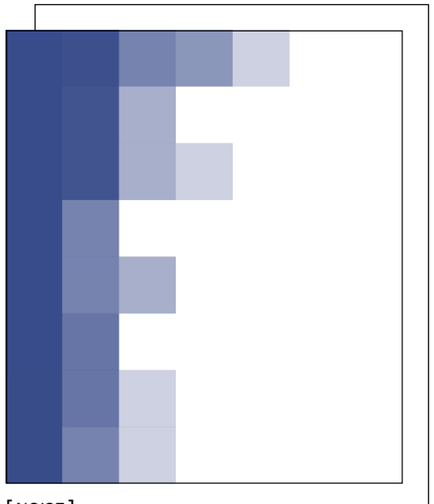


[18]

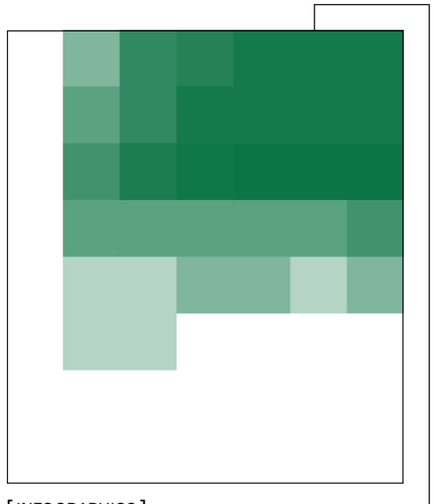


TOPIC AREAS

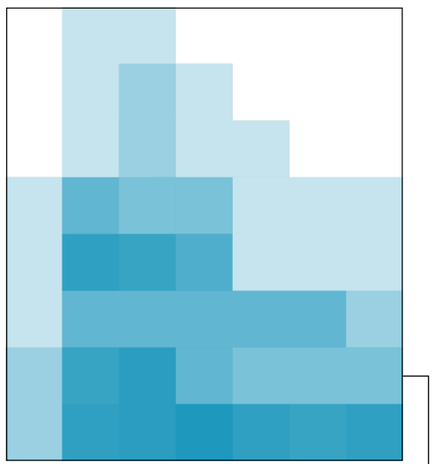
FINAL VISUALIZATION: The area belonging to each topic was isolated and overlayed to find the corrisponding coordinates of the XY plot values.



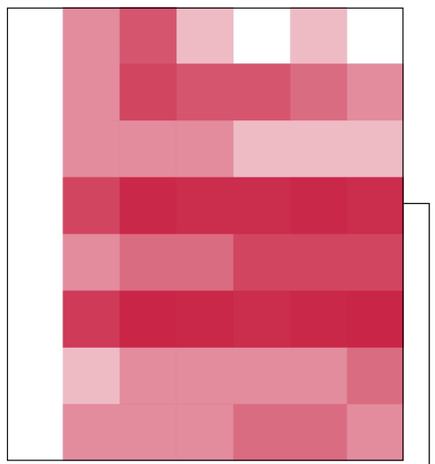
[NOISE]



[INFOGRAPHICS]

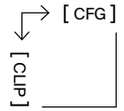


[URBAN LANDS]



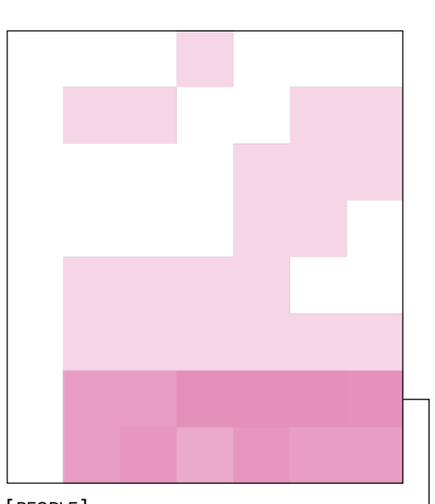
[PROTEST]

How to read it:

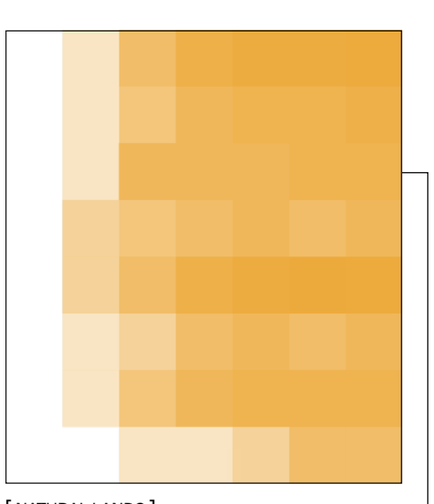


Probability to find the topic in the respective coordinate:

HIGH LOW



[PEOPLE]



[NATURAL LANDS]

[FINDING] ❶

Climate change related ***topics*** refer to ***specific coordinates*** of the plot.

Within the plots, we can clearly distinguish different coordinate values for each topic. The noise position is quite consistent in CFG: 1 column, whereas natural lands and protest are the topics that occupy the widest area. This means there is a higher probability of getting such thematic pictures in the image generation process. Infographics and people are less spread, but still significantly situated among the plot, being clearly positioned at the beginning and at the end of the creation process. Urban lands area has an average expansion but shows reiteration in CLIP 9 row.

TOPICS INTERPOLATION

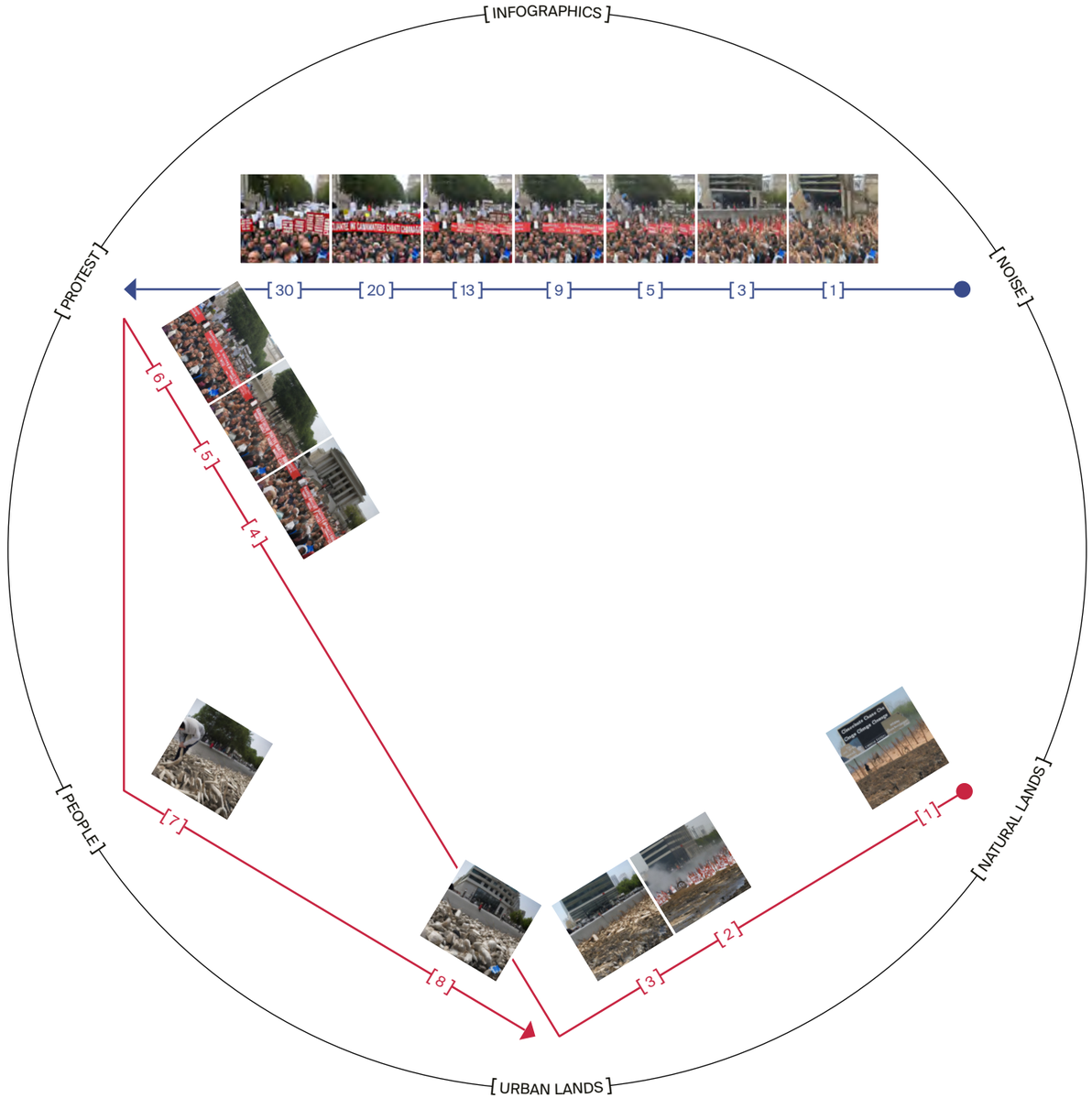
X/Y PLOT ANALYSIS:

The aim of this analysis is to find interpolations through topics inside each plot.

The selected column and row are where the biggest amount of shifts between topics occur.

How to read it:

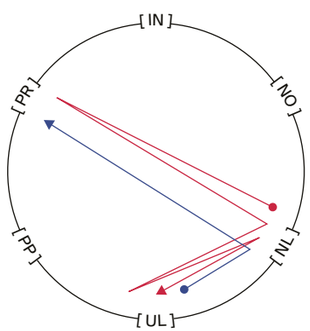
CFG gives a more gradual interpolation between two topics, while CLIP jumps from one topic to another.

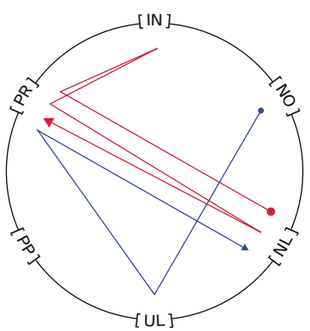


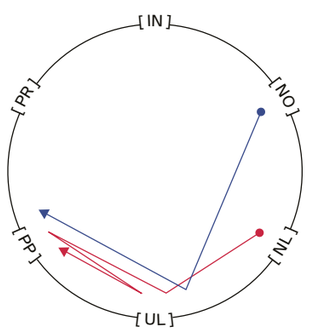
TOPICS INTERPOLATION

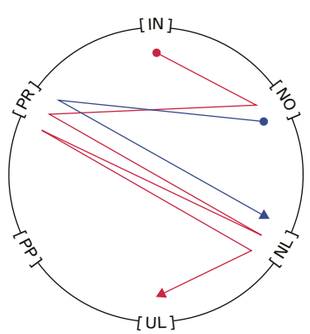
X/Y PLOT OVERVIEW:

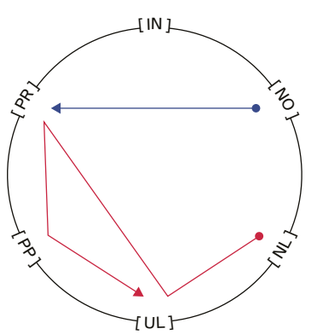
The analysis was conducted on all the plots to find patterns in the interpolation between topics.

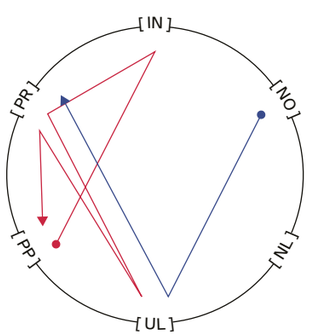
[1]


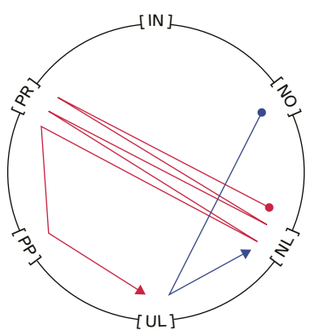
[2]


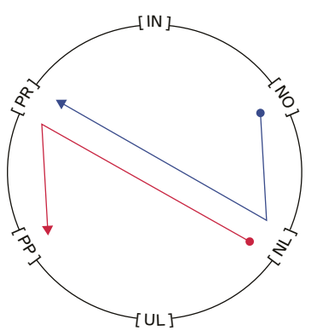
[3]


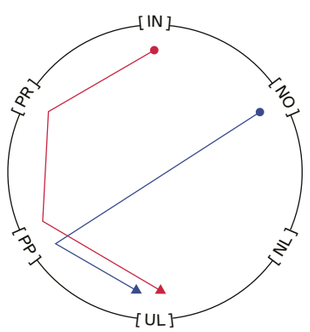
[4]


[5]


[6]


[7]


[8]


[9]


How to read it:

CLIP increment

CFG increment

[IN] NOISE

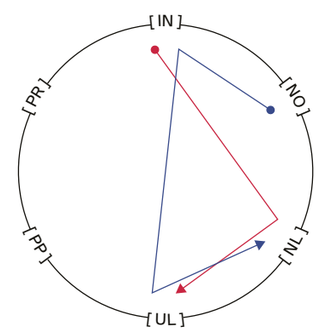
[PP] PEOPLE

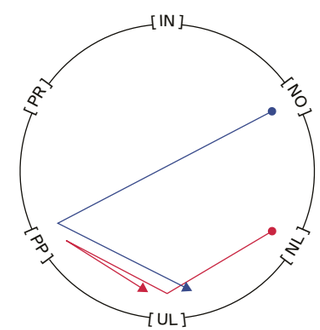
[UL] URBAN LANDS

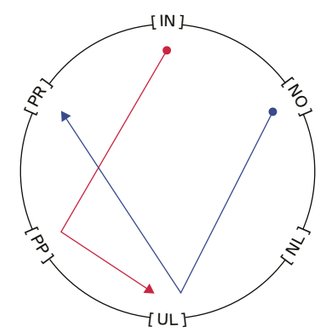
[NL] NATURAL LANDS

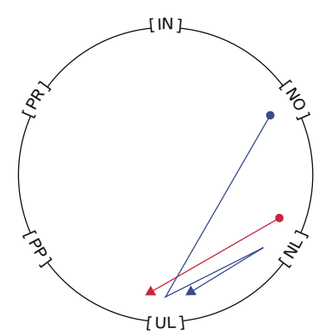
[PR] PROTEST

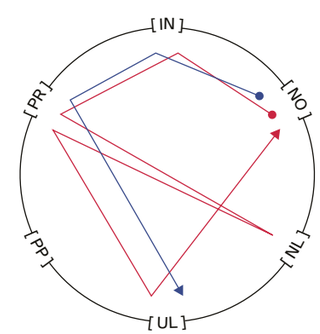
[IN] INFOGRAPHICS

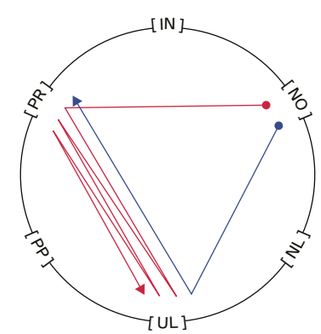
[10]


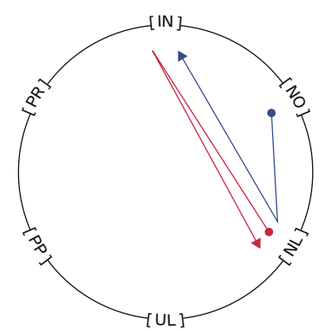
[11]


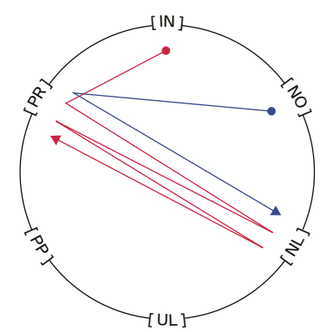
[12]


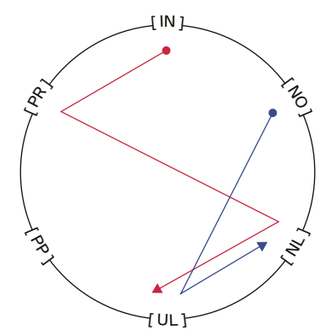
[13]


[14]


[15]


[16]


[17]


[18]


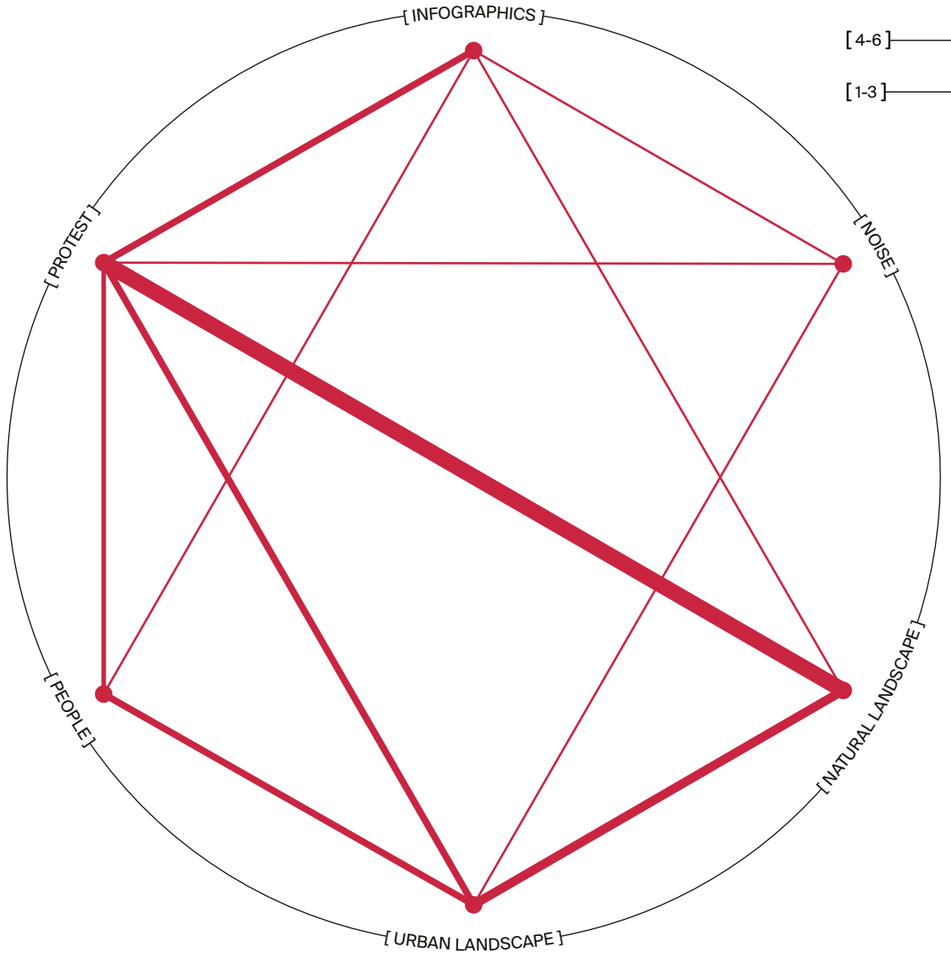
TOPICS INTERPOLATION

FINAL VISUALIZATION:

Topics that are more strongly linked are topics that are subject to frequent shifts between them.

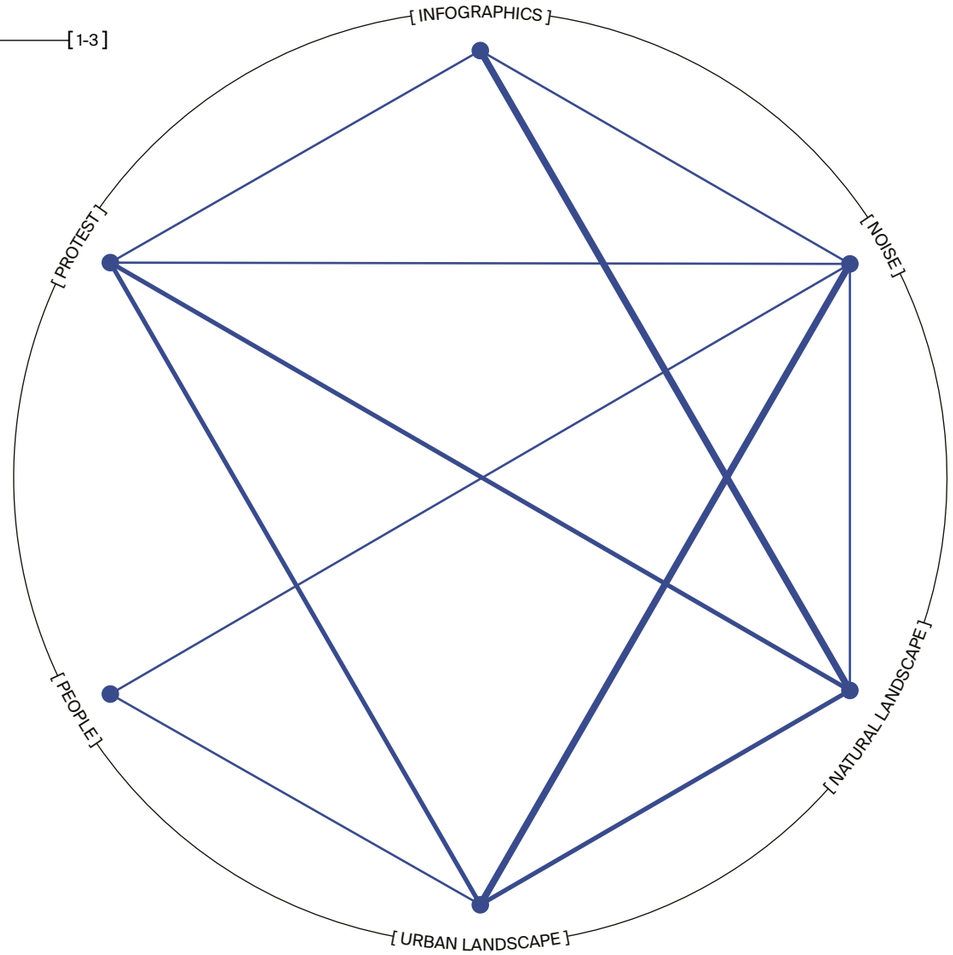
OF CONNECTIONS related to CLIP:

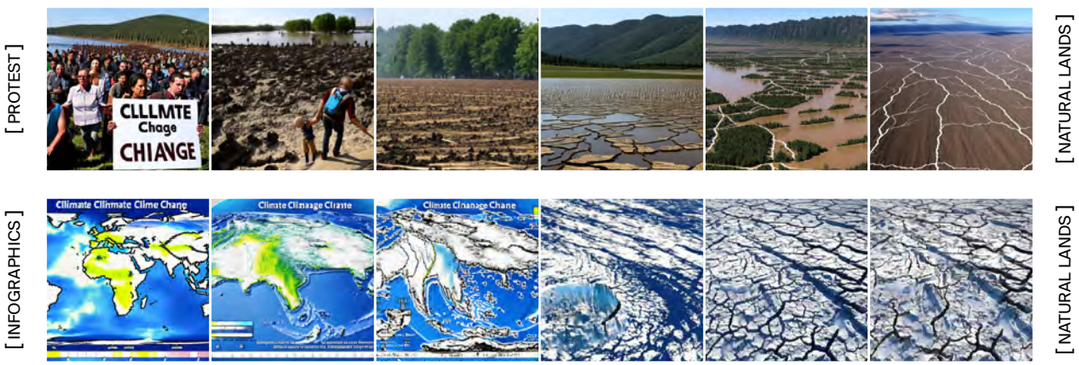
[19-21]
[10-12]
[7-9]
[4-6]
[1-3]



OF CONNECTIONS related to CFG:

[10-12]
[7-9]
[4-6]
[1-3]





[FINDING] ❷ ***Interpolation*** between topics is mainly about ***protests*** and ***natural landscapes.***

Interpolation between topics takes place more frequently between different CLIP values. Themes that are often associated with each other within CLIP increment are natural landscapes and protests, protests and infographics, as well as natural landscapes an urban landscapes. Within CFG increment topics that are often linked are mainly natural landscape and infographics, but also noise and urban landscapes.

RELEVANT IMAGES

X/Y PLOT ANALYSIS: This analysis aims at finding the coordinates within the plot of the most relevant generated images, as well as inspecting their content.

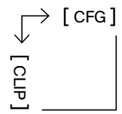
RELEVANCE CRITERIA: sharp, realistic, representative of the prompt, close to climate visuals principles.

RELEVANT IMAGES

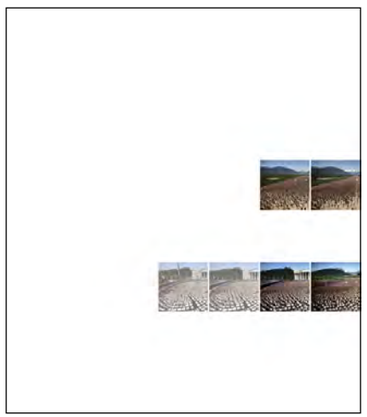
X/Y PLOT OVEWVIEW

The process was applied to all 18 plots and the most representative images were identified.

How to read it:



[1]

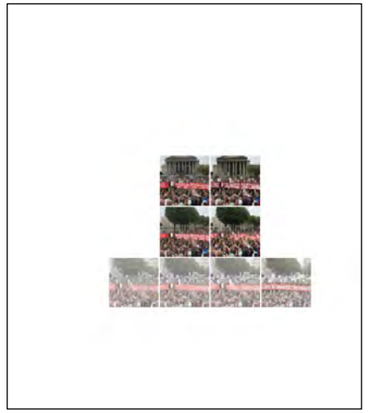


[2]

[3]

[4]

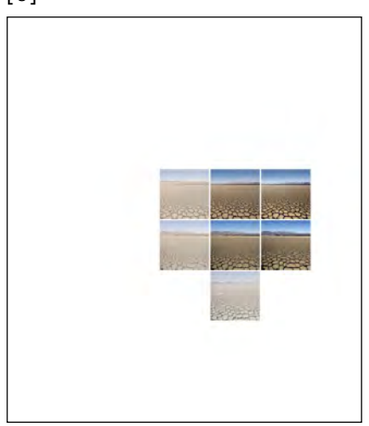
[5]



[6]

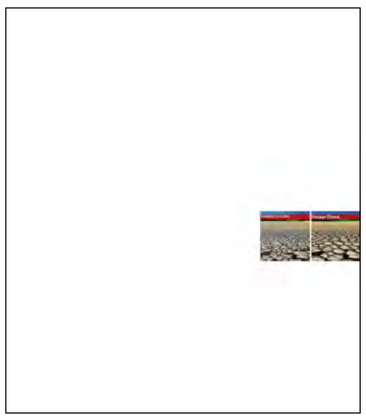
[7]

[8]



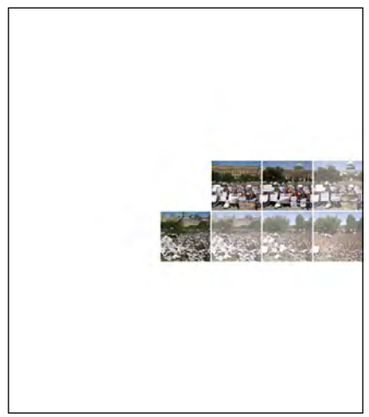
[9]

[10]

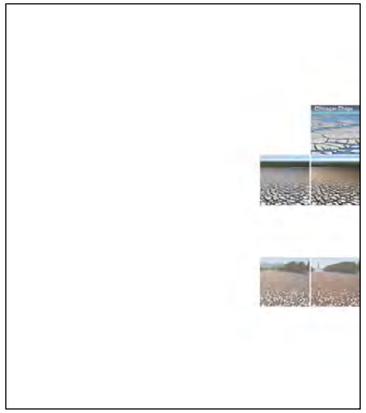


[11]

[12]



[13]



[14]

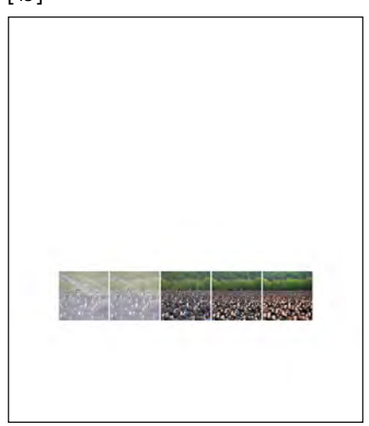
[15]

[16]

[NO RELEVANT IMAGES]

[17]

[18]



RELEVANT IMAGES

FINAL VISUALIZATION: All the relevant images were overlayed to find the area where they are more concentrated.

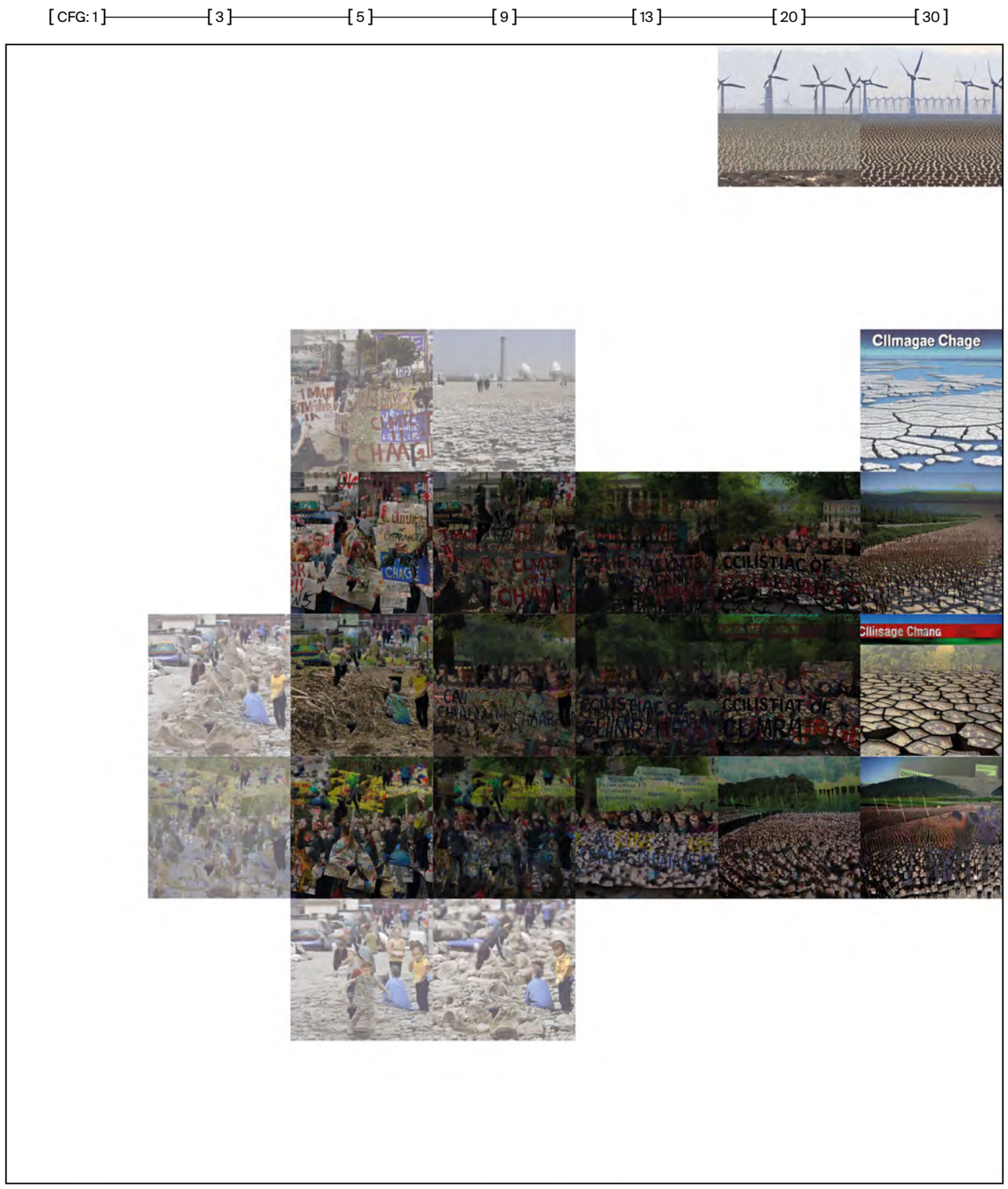


IMAGE CLUSTERS: All the relevant images selected through the chosen criteria are shown below and clustered in topics.

[FINDING] 3

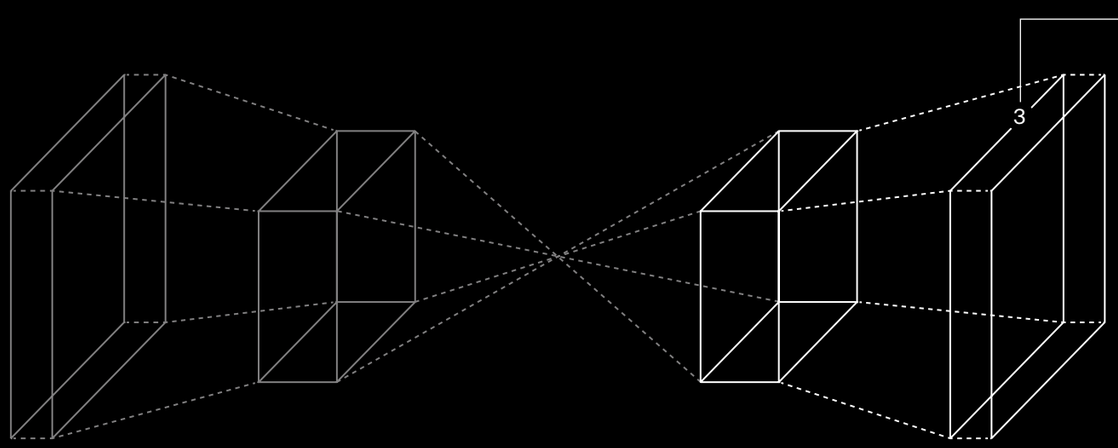
The most relevant images are more likely to be generated with ***default settings***

Increasing the values of CFG and CLIP can easily cause a loss in terms of image quality. Among all the images generated in the plots, the best ones came from the default settings given by Stable Diffusion.

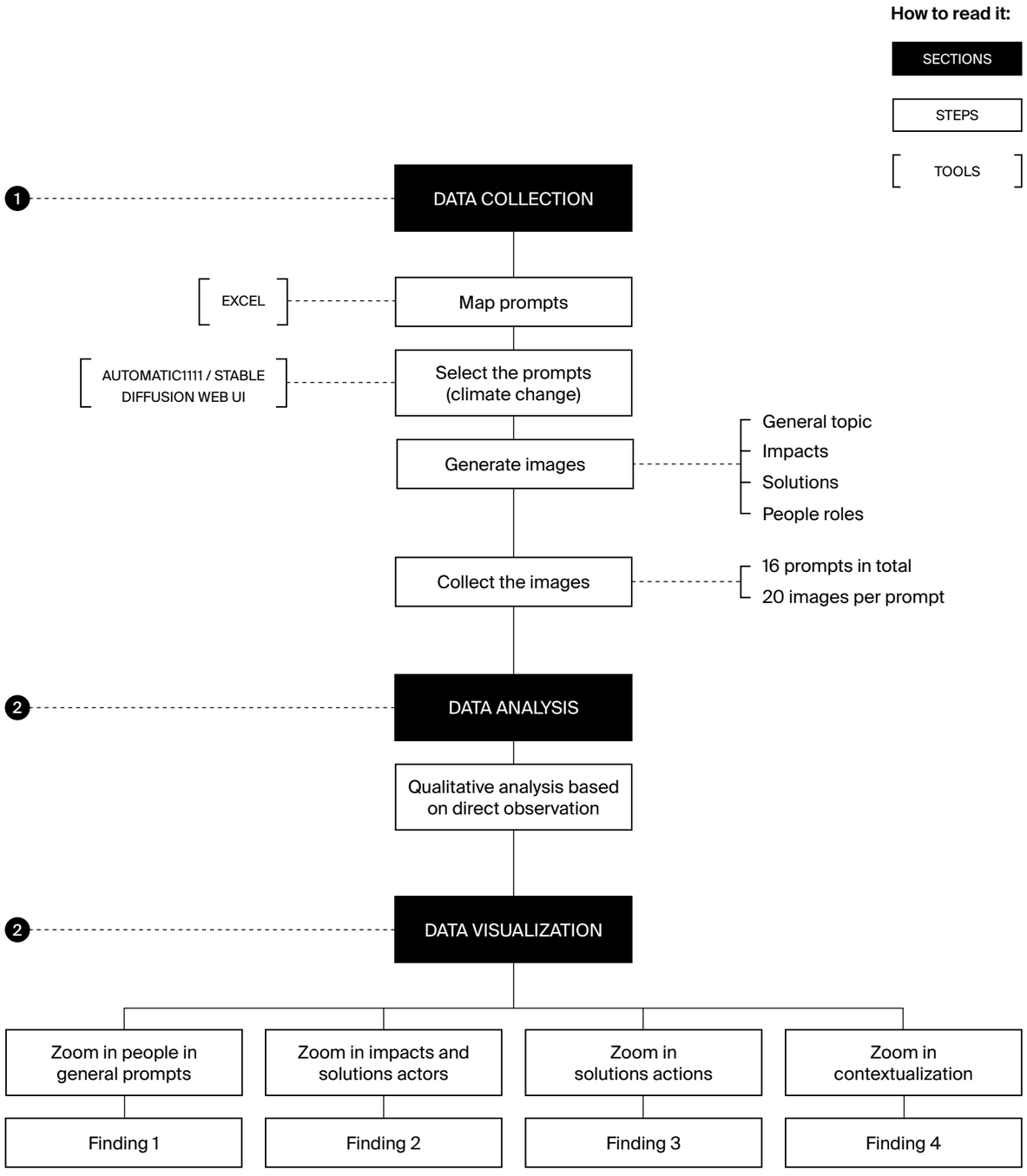
[FINDING] 4

The most relevant images are ***not consistent*** with ***climate change visual principles***

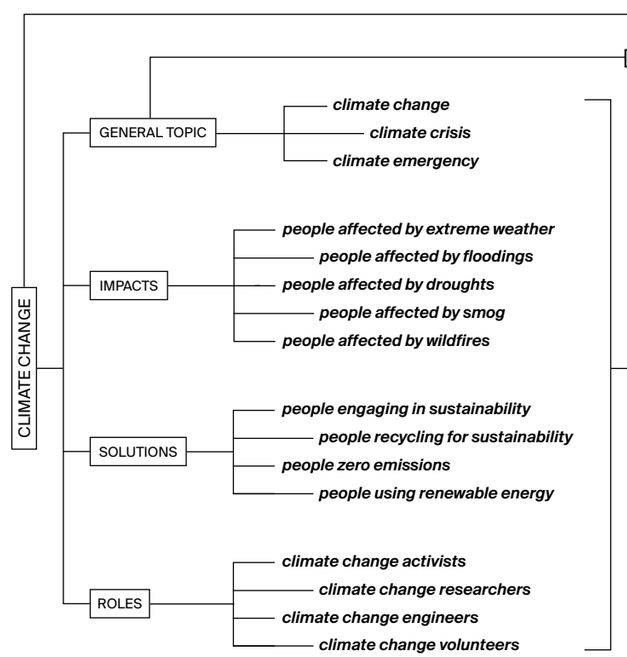
Among all the thousands of images generated, the most relevant ones are not able to successfully represent the topic of climate change in a meaningful way. The representation is highly stereotyped and repetitive over the same types of contents.



How does Stable Diffusion represent people and their actions related to climate change in generated images?



A *prompts map* was created starting from the topic and its subtopics.



The starting point is **Climate Change**.

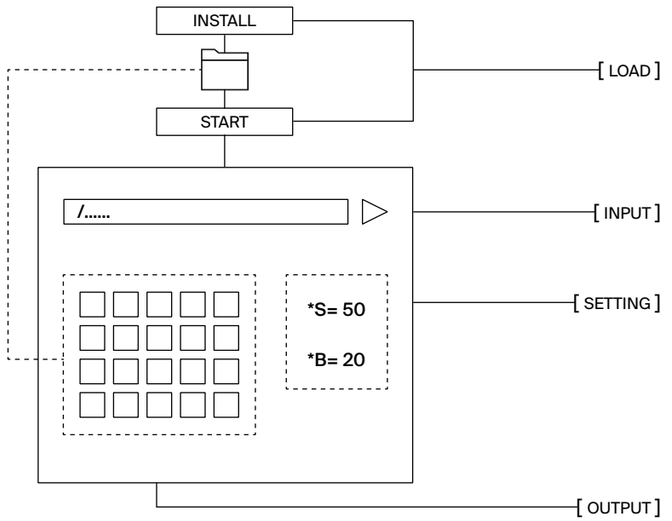
Different aspects of this topic were considered based on a research among organizations websites such as *UN.org*, resulting in 4 categories with **different levels of detail**.

4 categories

Different prompts have been created for each category in order to **find and analyse people** in the generated images, using both general and specific prompts with the keyword “people” forced.

16 prompts

AUTOMATIC1111 was used to generate batches of images all at once starting from each prompt.



Running **AUTOMATIC1111** from Google Colab will automatically **store the image set** generated by Stable Diffusion in Google Drive.

All the specific **prompts** of the prompts map have been tested.

The most important settings were adjusted to obtain a suitable sample of optimal images.

***S = Sampling steps**
The steps of the machine operation
***B = Batch Count**
The number of images to generate at once

20 images per prompt
320 images in total

GENERATED IMAGES OVERVIEW

OVERVIEW:

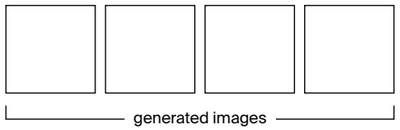
Each row presents the images genereted by the same prompt on the left. In this way we can instantly compare the results.

Settings:
Seed: random
Batch count: 20

How to read it:

[prompt category]

prompt used



[general topic]

climate change

climate crisis

[impacts]

climate emergency

people affected by extreme weather

people affected by floodings

people affected by droughts

people affected by smog

people affected by wildfires

GENERATED IMAGES OVERVIEW

OVERVIEW:

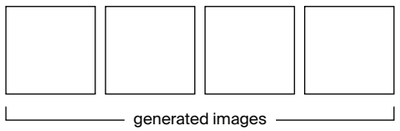
Each row presents the images genereted by the same prompt on the left. In this way we can instantly compare the results.

Settings:
Seed: random
Batch count: 20

How to read it:

[prompt category]

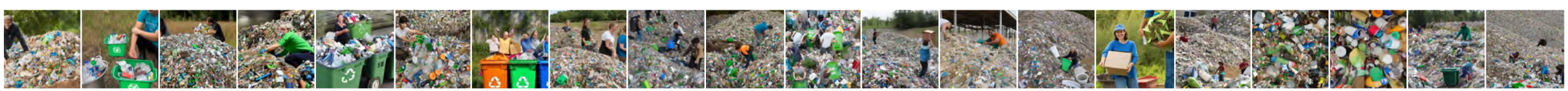
prompt used



[solutions]

people engaging in sustainability

people recycling for sustainability



people zero emissions

people using renewable energy

[people roles for solutions]

climate change activists

climate change researchers

climate change engineers

climate change volunteers

PEOPLE IN GENERAL PROMPTS

PEOPLE HIGHLIGHTING:

Highlighting only people let us see their frequency in images generated with a general prompt about climate change.

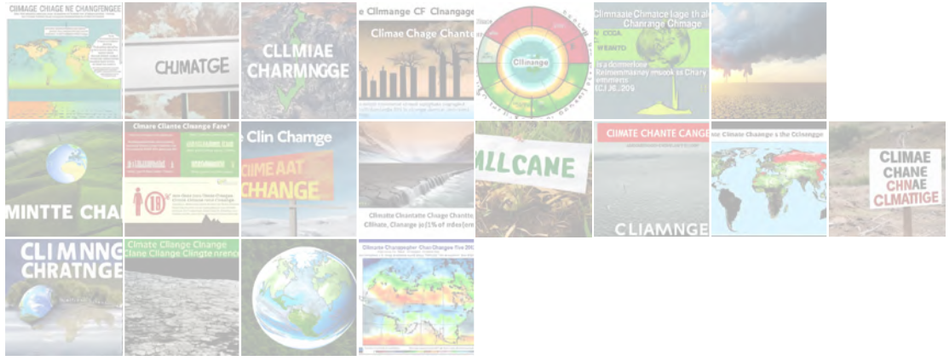
[prompt category]

prompt used

generated images

[general topic]

climate change



climate crisis

climate emergency

CLASSIFICATION BY SUBJECT:

Classifying these images is useful to understand in which kind of images and how often people appear.

[prompt category]

[CLASS]

classified images

[general topic]

[EARTH]

[LANDSCAPES]

[SIGNS]

[PROTESTS]

[TEXT/INFOGRAPHICS]

[FINDING] ❶

Most of the images that Stable Diffusion generates to represent the general topic of climate change ***doesn't contain people***.

If images are generated with a non-specific prompt, the AI creates outputs with almost no people. People start to compare when writing prompts closer to human themes such as *crisis* or *emergency*. Classifying the outputs let us see that all the people are represented in a protesting situation, but those images are anyway a lower percentage than images with text inside (despite its inability to write meaningful things) or unreal landscapes images.

IMPACTS AND SOLUTIONS ACTORS

COMPARISON AND HIGHLIGHTING ANALYSIS: The most relevant images of each prompt related to impacts and solutions are compared, highlighting actors and context.

[impacts]

people affected by extreme weather

people affected by floodings

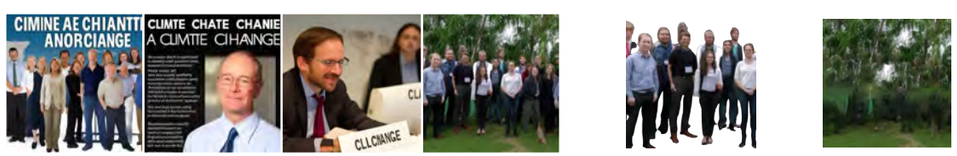
people affected by droughts

people affected by smog

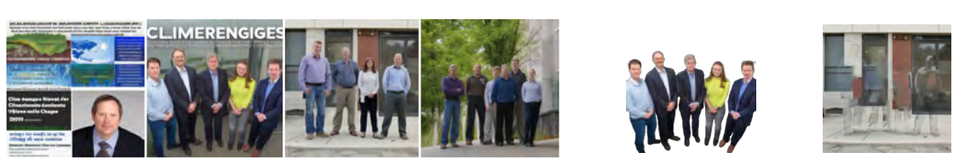
[people roles for solutions]

climate change activists

climate change researchers



climate change engineers



climate change volunteers

How to read it:

[prompt category]

prompt used

[FINDING] ❷

There are visible ***ethnicity-based*** and ***gender-based biases*** when asked to represent ***victims*** and ***helpers***.

The algorithm represents impacts' and solutions' actors with very different ethnicities and contexts showing a really stereotyped view also for gender. Specifically, people affected by climate change have always asian or african features and are in extreme environment or disaster situations, with a lot of women/families. On the contrary, people finding solutions or acting against climate change have always western features, with a prevalence of men in suits for reaserchers and engineers and women as volunteers, mainly in urban/parks settings.

SOLUTIONS ACTIONS

ACTIONS AND COLORS ANALYSIS:

The images about solutions are compared and the main colors and actions are highlighted in order to underline stereotypes.

[solutions]

people engaging in sustainability

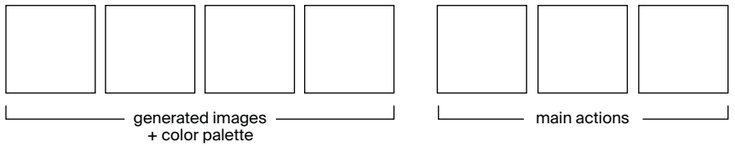
people recycling for sustainability

people zero emissions

How to read it:

[prompt category]

prompt used



people engaging in sustainability

people recycling for sustainability

people zero emissions

[FINDING] ❸

Concepts related to ***climate change solutions*** are always represented with the same ***banalized*** and ***stereotyped actions***.

By using extended themes such as sustainability, the algorithm fails to represent all the possible variations creating stereotyped and trivialised images as output. The palettes mainly based on green stresses the stereotype even more. The actions are repeated and they're often meaningless or bordering on the absurd.

CONTEXTUALIZATION

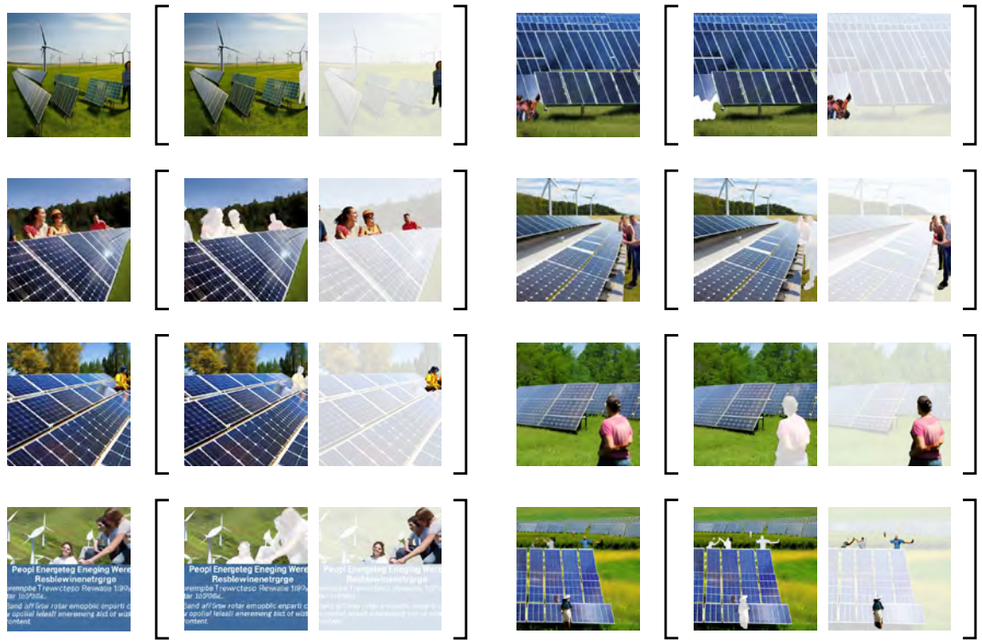
HIGHLIGHTING ANALYSIS: The most relevant images were selected to isolate people and context and underline the lack of contextualization.

[impacts]

people affected by wildfires

[solutions]

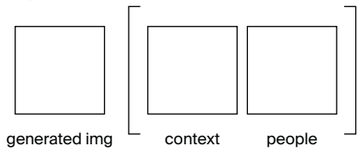
people using renewable energy

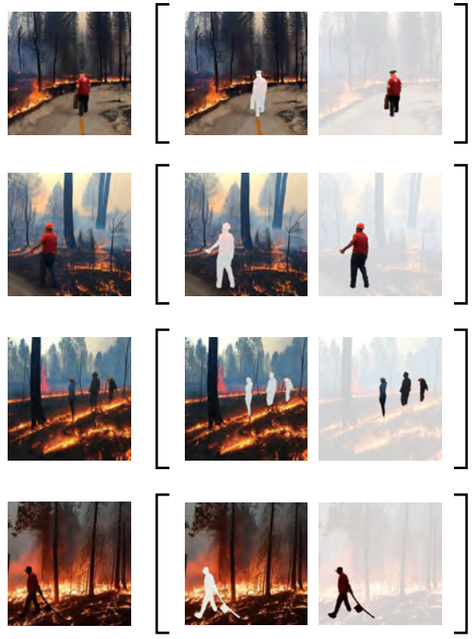


How to read it:

[prompt category]

prompt used





[FINDING] ❹

Often people ***can't be generated*** in ***contextualized situations***, resulting in an ***unrealistic collage***.

Separating the subjects and the context of some of the images generated underlines clearly how there's no correlation between them. The machine isn't able to generate contextualized situations, resulting in people completely out of context just pasted on a background. Also the type of context of these images is always the same, banalizing the concepts.

[COVER IMAGE PROMPT]

Incredible report about Stable Diffusion climate change visuals trending on Final Synthesis Studio