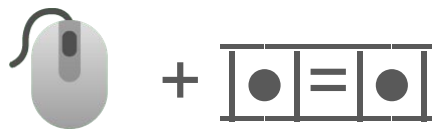


final

will



Please make ***the focus of your presentation on your final results***. The instructors are now familiar with your methods and motivations. Ensure that you leave yourself enough time to show your results in full. Many of the project 1-3 presentations spent a lot of time talking about experiments, methods, and motivations and did not have enough time to show results. Please don't let this happen for your final project. ***You may even want to start your presentation with the results***, and if you have extra time work your way back and explain the methods.

Please make ***the focus of your presentation on your final results [...]*** You may even want to start your presentation with the results:

I've **ripped a seismic hole in established data preparation for text to image models**, yielding results that demand further research into how **ignoring grammar for fear of data inconsistency can distort a model's internal logic** governing what constitutes coherent text, coherent image, and their relationship transcendent of human comprehension

S

si

sin

sing

singl

single

single c

single ch

single cha

single char

single chara

single charac

single charact

single character

single character

single character i

single character in

single character inp

single character input

single character input

single character inputs

single character inputs

Will Rinkoff

““““

Please make **the focus of your presentation on your final results**. The instructors are now familiar with your methods and motivations. Ensure that you leave yourself enough time to show your results in full. Many of the project 1-3 presentations spent a lot of time talking about experiments, methods, and motivations and did not have enough time to show results. Please don't let this happen for your final project. **You may even want to start your presentation with the results**, and if you have extra time work your way back and explain the methods.

””””

too late :/

Inspiration

this  >

- ~~impressive~~
- ~~contribution over the baseline/
original approach was significant~~
- ~~provokes meaningful questions~~
- ~~or imaginations~~ <--
- ~~kinda fun~~
- surprising



New Domain: dalle-mini

Differences:

- text -> 256x256 images (not text -> text)
 - trained on image-caption pairs
- trained on smaller dataset (compared to gpt 2)
 - drew from “Conceptual Captions” + “Conceptual 12M” + “the OpenAI subset of YFCC100M” datasets
 - 15 mil. for seq2seq component + 2 mil. tuning vqgan encoder
- standard ascii characters instead of emoji
 - “#”, “~”, “.”, “*”, etc...
 - no “passport control” emoji
 - 

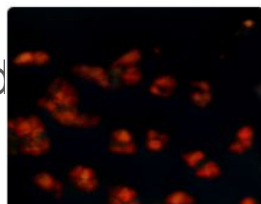
The Means

“Copy of Copy of DALL·E mini - Inference pipeline.ipynb”

- two modifications
 - “queue multiple inputs”
 - “save images as files”

Initial Results

What d



17.58.png



19.53.png



19.73.png



20.01.png



20.38.png



20.53.png



20.74.png



20.88.png



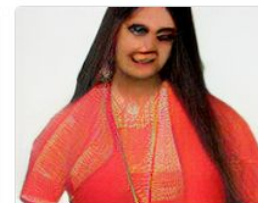
20.90.png



20.96.png



21.09.png



21.14.png



21.22.png



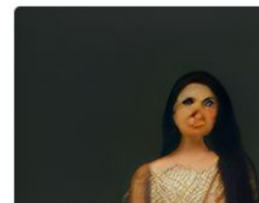
21.42.png



21.46.png



21.59.png



21.68.png



22.04.png



{_18.44.png



{_18.61.png



{_18.73.png



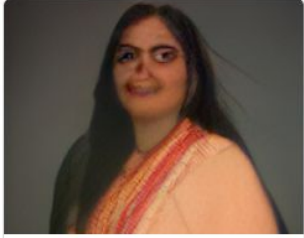
{_19.47.png



{_19.70.png



{_19.82.png



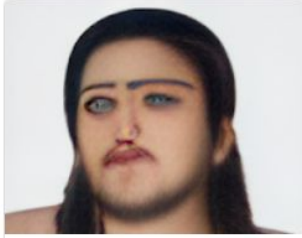
{_19.92.png



{_20.28.png



{_20.29.png



{_20.30.png



{_20.85.png



{_20.88.png



{_20.91.png



{_20.95.png



{_21.23.png



{_21.24.png



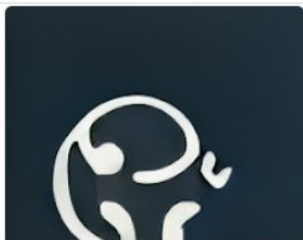
{_22.13.png



{_22.38.png



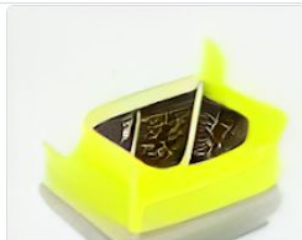
 %_15.38.png



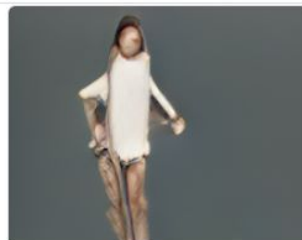
 %_16.50.png



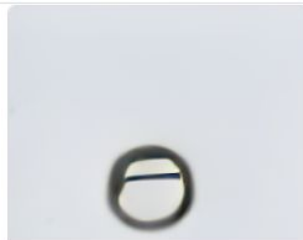
 %_18.29.png



 %_19.04.png



 %_19.47.png



 %_19.76.png



 %_20.06.png



 %_20.21.png



 %_20.36.png



 %_20.45.png



 %_20.59.png



 %_20.71.png



 %_20.88.png



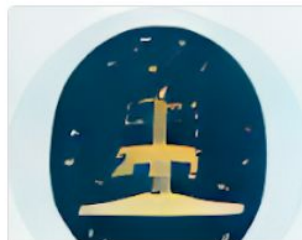
 %_20.93.png



 %_20.95.png



 %_21.09.png



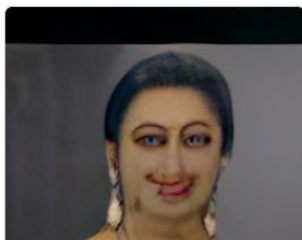
 %_21.10.png



 %_21.17.png



 _13.92.png



 _16.08.png



 _16.71.png



 _16.79.png



 _17.28.png



 _17.65.png



 _17.90.png



 _18.33.png



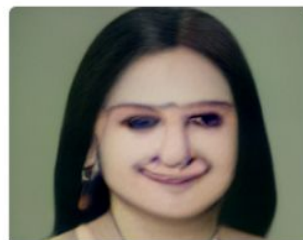
 _18.40.png



 _18.51.png



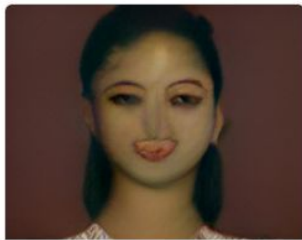
 _18.52.png



 _18.55.png



 _18.59.png



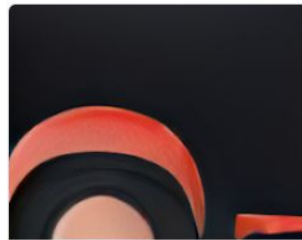
 _18.77.png



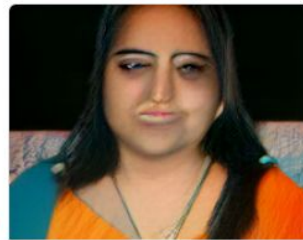
 _19.04.png



 _19.09.png

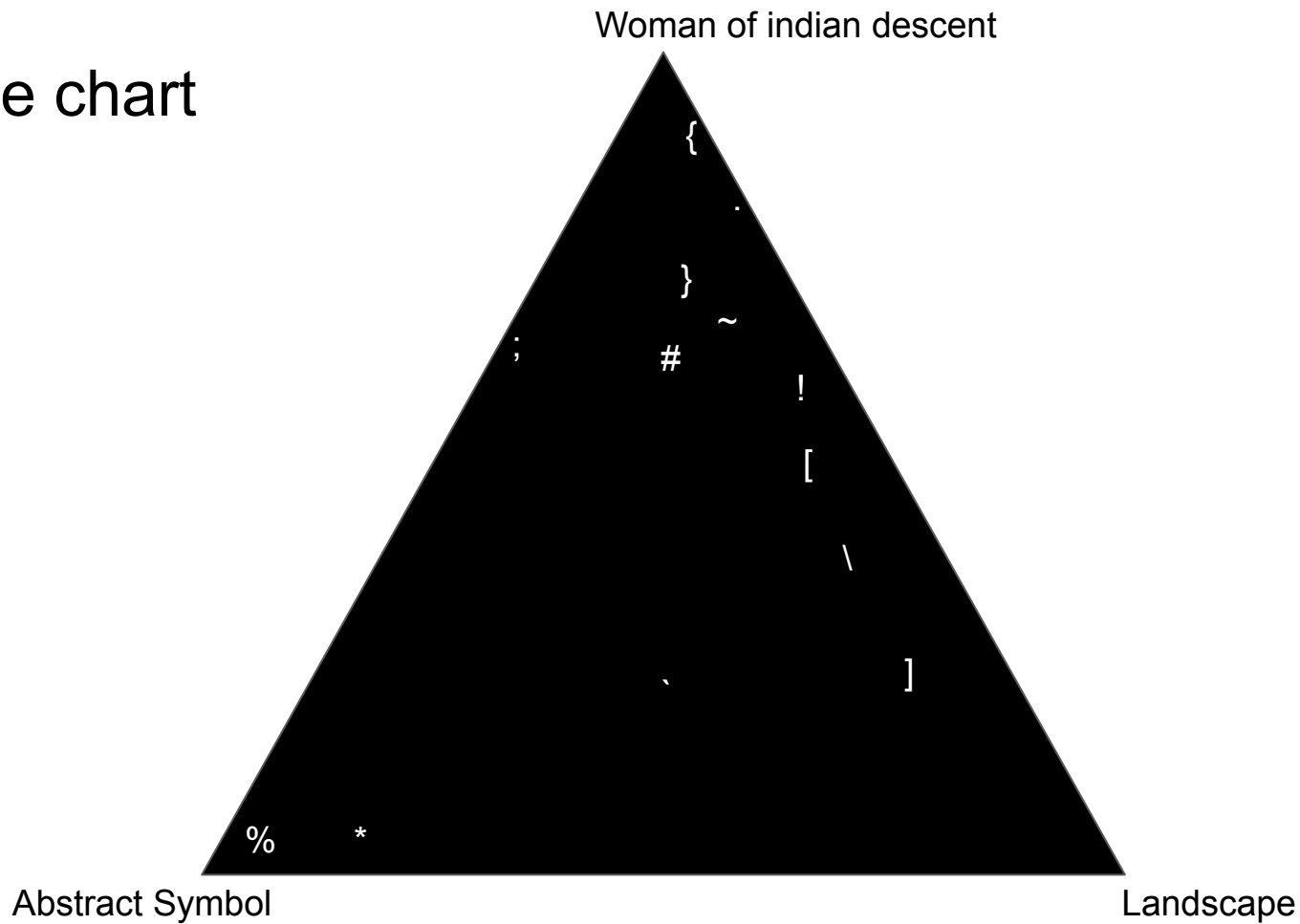


 _19.14.png



 _20.01.png

triangle chart



Thoughts

- huh
- what's in those datasets?

Conceptual Captions Dataset quirks

“””

The following steps are applied to achieve text transformations:

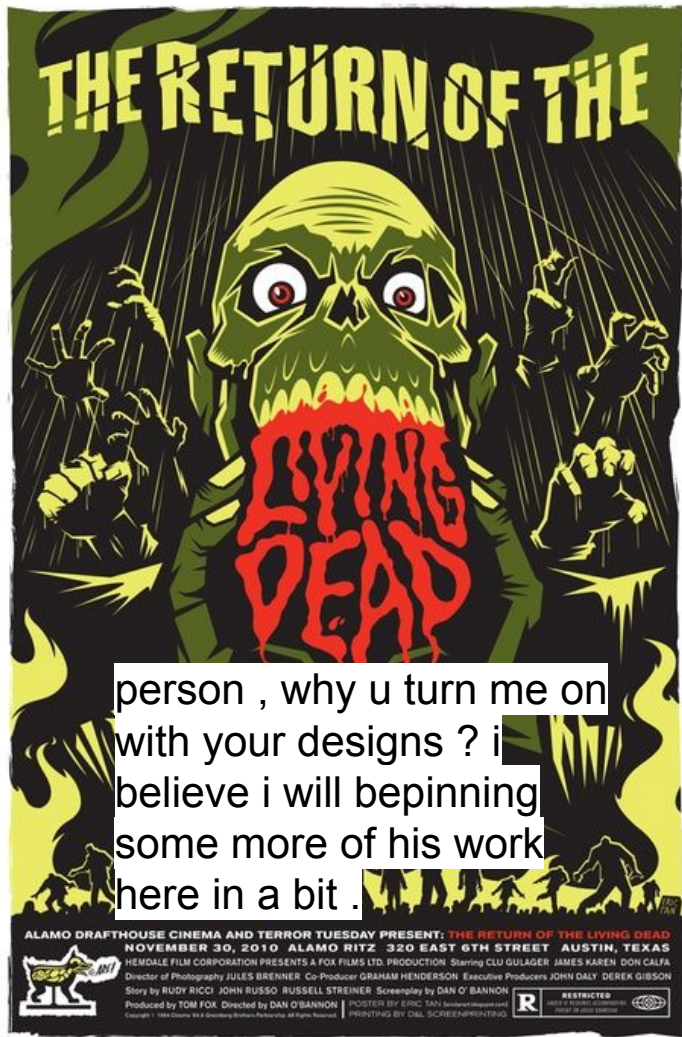
- noun modifiers of certain types (proper nouns, numbers, units) are removed;
- dates, durations, and preposition-based locations (e.g., “in Los Angeles”) are removed;
- named-entities are identified, matched against the KG entries, and substitute with their hypernym;
- resulting coordination noun-phrases with the same head (e.g., “actor and actor”) are resolved into a single-head, pluralized form (e.g., “actors”);

“”” - CC paper (<https://aclanthology.org/P18-1238.pdf>)



... and local people

[s://aclanthology.org/P](https://aclanthology.org/P)



Observation

punct	^	∨	☒ Highlight <u>A</u> ll	☐ Match <u>C</u> ase	☐ Match D <u>i</u> acritics	☐ <u>W</u> hole Words	Phrase not found
period	^	∨	☒ Highlight <u>A</u> ll	☐ Match <u>C</u> ase	☐ Match D <u>i</u> acritics	☐ <u>W</u> hole Words	Phrase not found
comma	^	∨	☒ Highlight <u>A</u> ll	☐ Match <u>C</u> ase	☐ Match D <u>i</u> acritics	☐ <u>W</u> hole Words	Phrase not found
grammar	^	∨	☒ Highlight <u>A</u> ll	☐ Match <u>C</u> ase	☐ Match D <u>i</u> acritics	☐ <u>W</u> hole Words	Phrase not found

Conceptual 12M dataset quirks

“The only exception to the “keep alt-texts as raw as possible” rule is performing person-name substitutions, which we identify as necessary to protect the privacy of the individuals in these images. For this step, we use the Google Cloud Natural Language APIs to detect all named entities of type Person, and substitute them by a special token <PERSON>. Around 25% of all the alt-texts in CC12M are transformed in this fashion.”

- 12M paper” (<https://arxiv.org/abs/2102.08981>)



The source of Anime quotes & Manga quotes : Photo
<PERSON>, Manga Quotes, Art Images, Fan Art, Thoughts,
Think, Anime, Crying, Random

Observation

punct ^ v ☒ Highlight All ☐ Match Case ☐ Match Diacritics ☐ Whole Words Phrase not found

comma ^ v ☒ Highlight All ☐ Match Case ☐ Match Diacritics ☐ Whole Words Phrase not found

grammar ^ v ☒ Highlight All ☐ Match Case ☐ Match Diacritics ☐ Whole Words Phrase not found

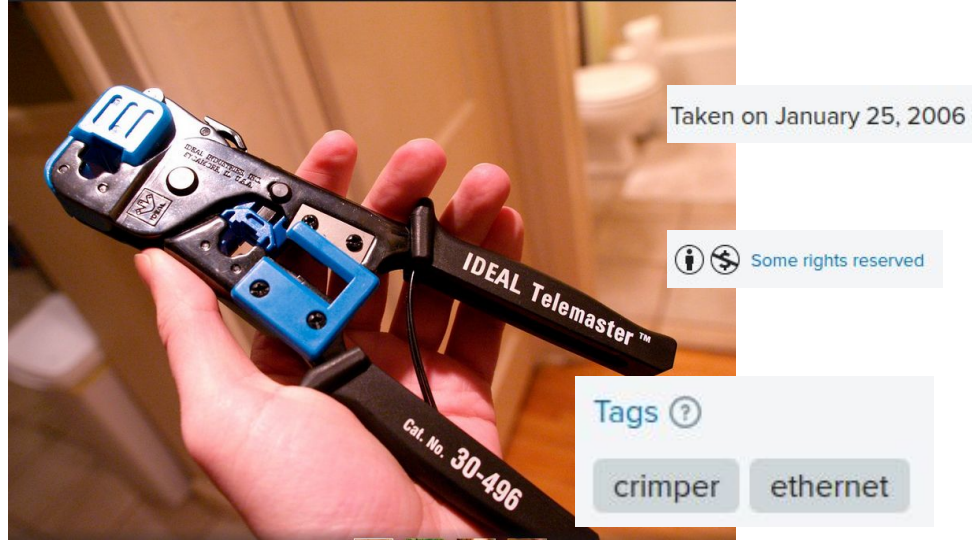
period ^ v ☒ Highlight All ☐ Match Case ☐ Match Diacritics ☐ Whole Words Phrase not found

YFCC100m dataset quirks

No single quote about text processing
because they just use flickr metadata
(photo tags, timestamps, geotags, and
cameras, licence, etc...)

- yfcc100m

paper



.	images & videos	Search
~	images & videos	Search
#	images & videos	Search

Results: 0 out of 100,000,000 items

 Download Results

[illegible]

grammar ^ v ☒ Highlight All ☐ Match Case ☐ Match Diacritics ☐ Whole Words Phrase not found

comma ^ v ☒ Highlight All ☐ Match Case ☐ Match Diacritics ☐ Whole Words Phrase not found

punc ^ v ☒ Highlight All ☐ Match Case ☐ Match Diacritics ☐ Whole Words Phrase not found

period ^ v ☒ Highlight All ☐ Match Case ☐ Match Diacritics ☐ Whole Words 1 of 4 matches

time period ^ v ☒ Highlight All ☐ Match Case ☐ Match Diacritics ☐ Whole Words 1 of 2 matches

the period ^ v ☒ Highlight All ☐ Match Case ☐ Match Diacritics ☐ Whole Words 1 of 2 matches

Future Work

- Dive further into datasets used
 - understand what may have led to the consistency
- Investigate text to image models trained with standardized grammar
- See if output for single character inputs are categorically similar across different models
 - ie. generated images revolve around a handful of themes
- Try other sets of unfamiliar single characters
 - ie. arabic alphabet, escape codes, chess symbols
 - □□□□□□ <- they're all here
- Explore model analysis/visualization tools to compare how familiar multi-word inputs diverge from unfamiliar single-character inputs

THANK

for listening

you

