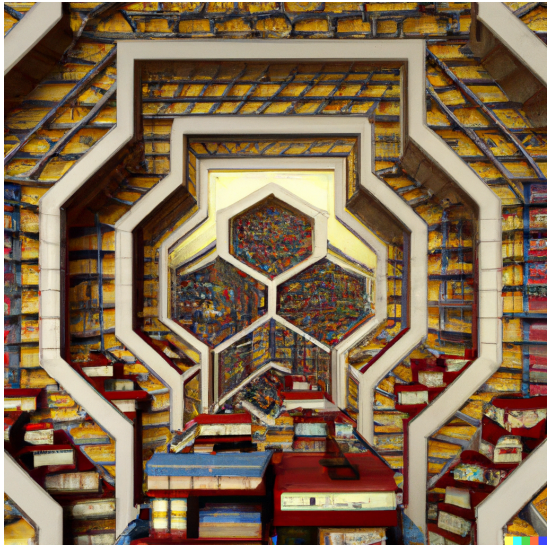
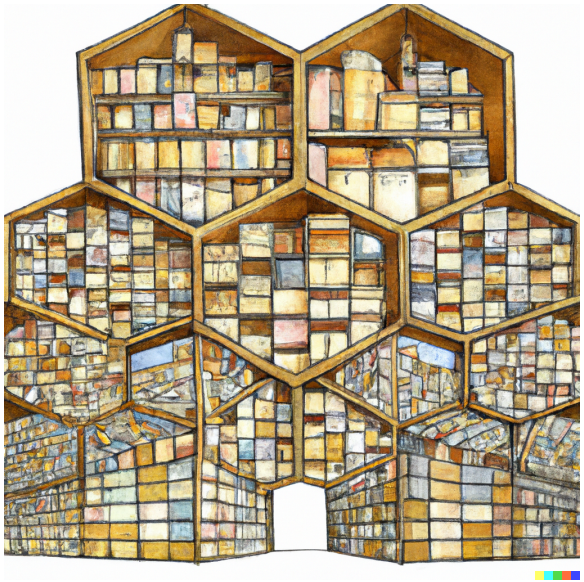




Natural Language Processing

Info 159/259

Lecture 24: Multimodal NLP (April 25, 2023)

David Bamman, UC Berkeley

NLP subfield survey — due May 9

- 2000-word survey for a specific NLP subfield of your choice (e.g., coreference resolution, question answering, interpretability, narrative generation, etc.), synthesizing at least 25 papers published at ACL, EMNLP, NAACL, EACL, AACL, *Transactions of the ACL* or *Computational Linguistics* (see bCourses for official details).

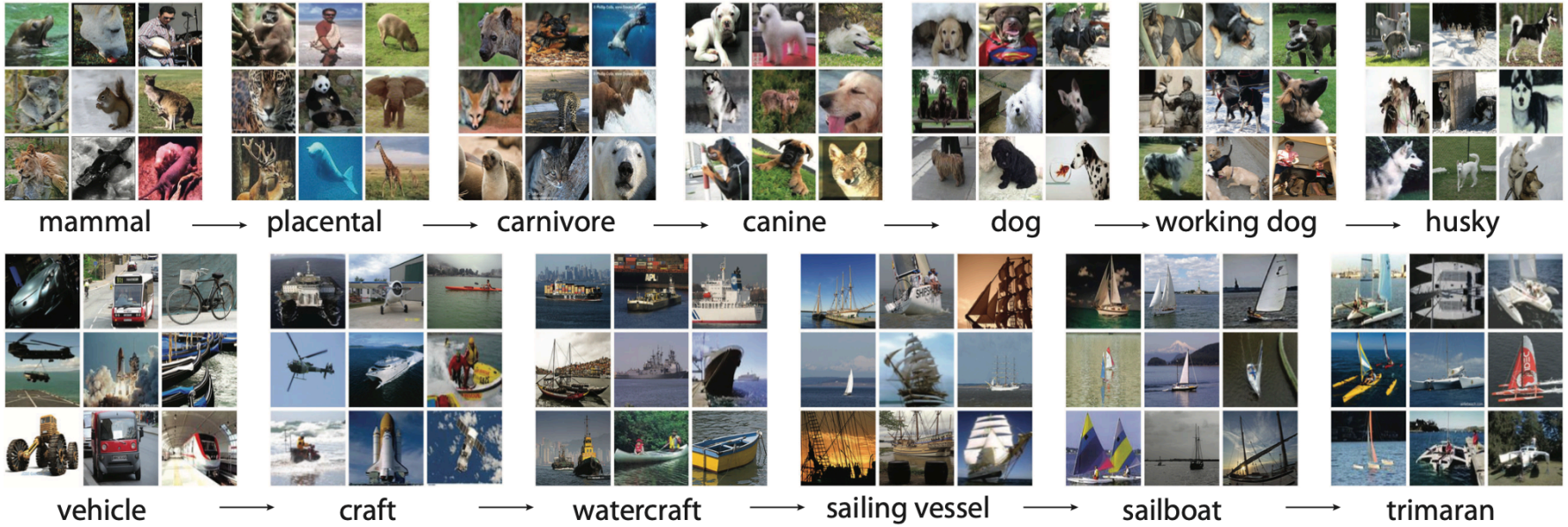
- This survey should be able to provide a newcomer (such as yourself at the start of the semester) a sense of the current state of the art in that subfield in 2023, the major historical papers that have defined that area, and the different schools of thought within it.

- No slip days! See bCourses assignment for policy on late assignments and list of example venues.

Info 259
Project presentations

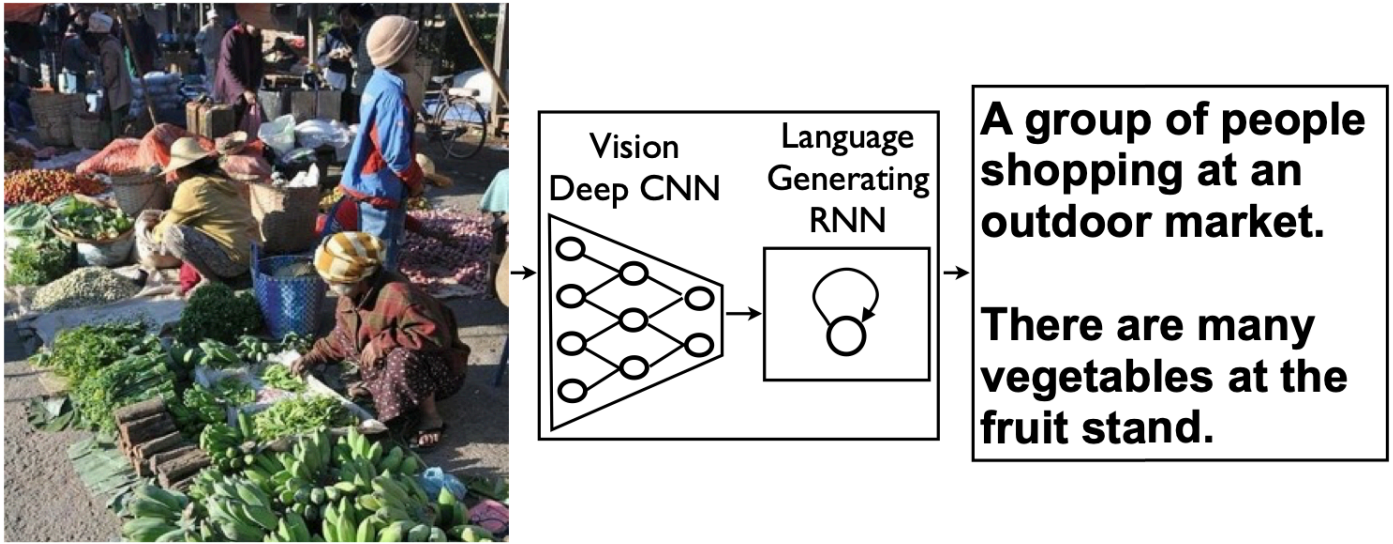
- 2-3:30pm Thursday 4/27 on Zoom.
- Prepare a 5-minute presentation of your project to present to the class; be prepared to take questions from the audience.
- The project presentations **won't** be recorded.

ImageNet



Deng et al. 2009, “ImageNet: A Large-Scale Hierarchical Image Database”

Image Captioning

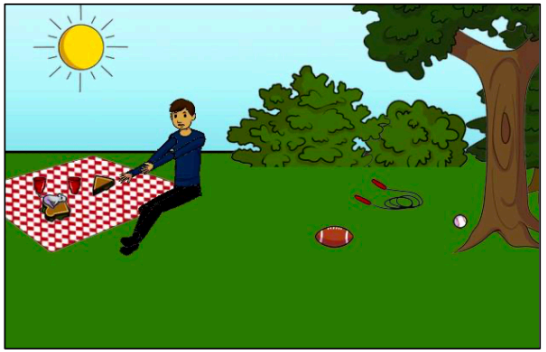


Vinyals et al. 2015, “Show and Tell: A Neural Image Caption Generator”

Visual QA

What color are her eyes?
What is the mustache made of?

How many slices of pizza are there?
Is this a vegetarian pizza?

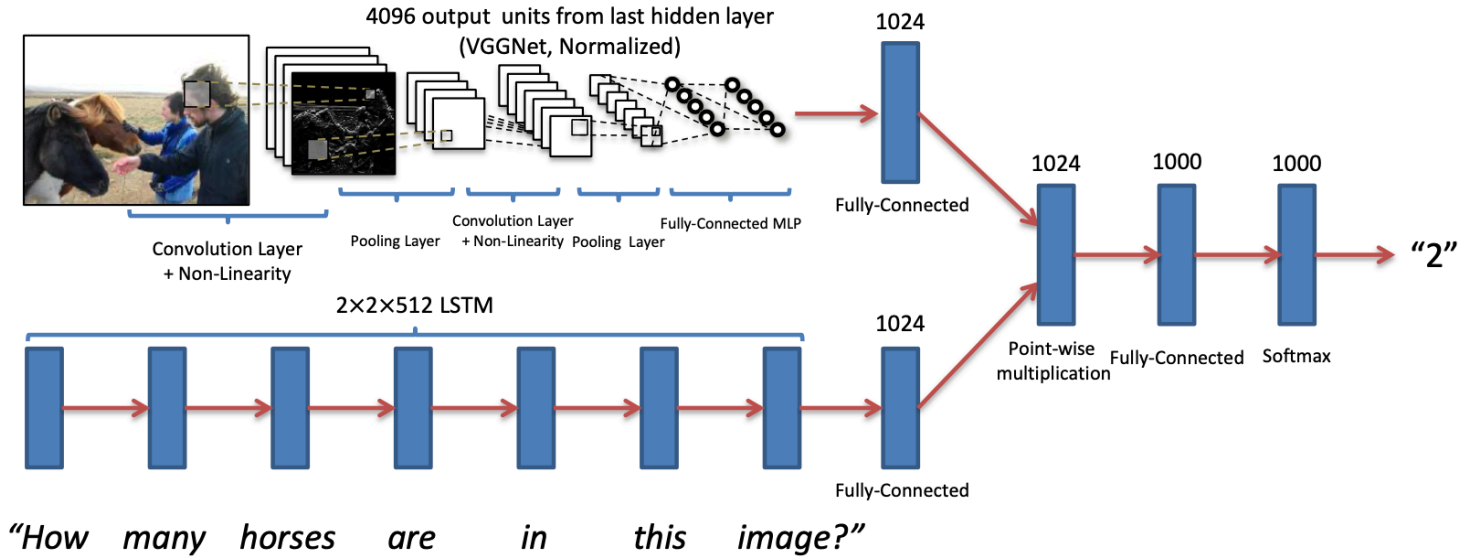


Is this person expecting company?
What is just under the tree?

Does it appear to be rainy?
Does this person have 20/20 vision?

Agarwal et al. 2016, “VQA: Visual Question Answering”

Visual QA



Agarwal et al. 2016, "VQA: Visual Question Answering"

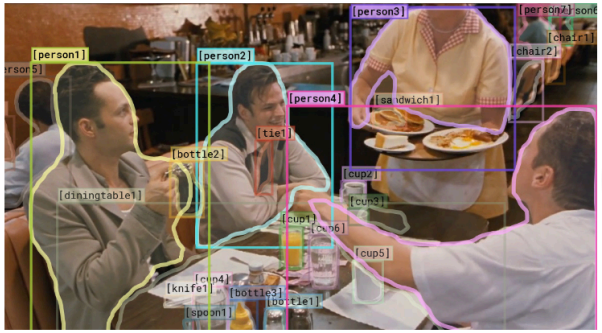
Visual Reasoning

The left image contains twice the number of dogs as the right image, and at least two dogs in total are standing.

One image shows exactly two brown acorns in back-to-back caps on green foliage.

Suhr et al. 2018, “A Corpus for Reasoning About Natural Language Grounded in Photographs”

Visual Commonsense Reasoning

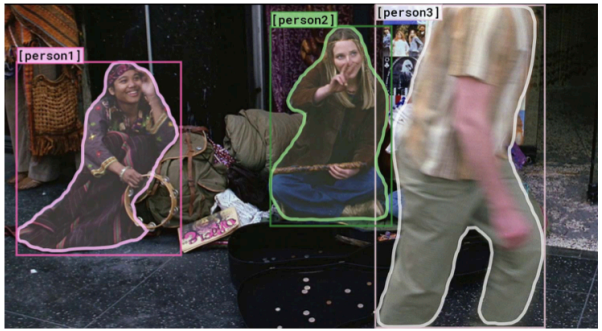


Why is [person4] pointing at [person1]?

a) **He is telling [person3] that [person1] ordered the pancakes.**
b) He just told a joke.
c) He is feeling accusatory towards [person1].
d) He is giving [person1] directions.

*I chose **a)** because...*

a) [person1] has the pancakes in front of him.
b) [person4] is taking everyone's order and asked for clarification.
c) [person3] is looking at the pancakes and both she and [person2] are smiling slightly.
d) **[person3] is delivering food to the table, and she might not know whose order is whose.**



How did [person2] get the money that's in front of her?

a) [person2] is selling things on the street.
b) **[person2] earned this money playing music.**
c) She may work jobs for the mafia.
d) She won money playing poker.

*I chose **b)** because...*

a) She is playing guitar for money.
b) [person2] is a professional musician in an orchestra.
c) **[person2] and [person1] are both holding instruments, and were probably busking for that money.**
d) [person1] is putting money in [person2]'s tip jar, while she plays music.

Zellers et al. 2018, "From Recognition to Cognition: Visual Commonsense Reasoning"

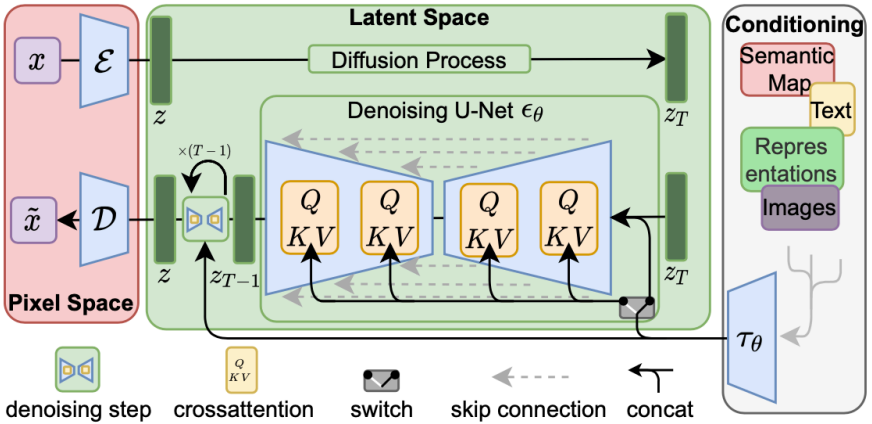
Text-to-Image Generation

- DALL-E (Ramesh et al. 2022)
- Transformer over streams of text+image tokens trained on Conceptual Captions (3.3 M text/ image pairs)

(a) a tapir made of accordion. a tapir with the texture of an accordion.

Stable Diffusion

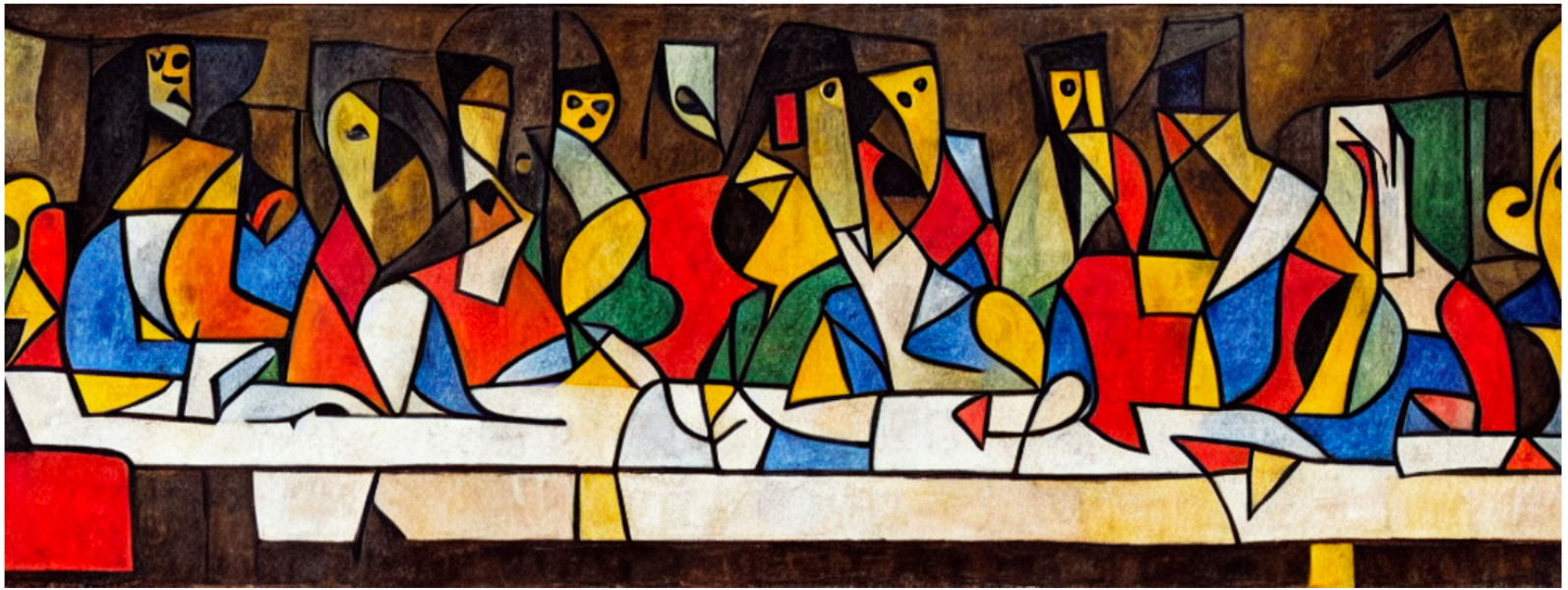
- Train image autoencoder to generate lower-dimensional perceptual representation space
- Learn diffusion model on the lower-dimensional latent space



- Scales to higher-dimensional data (i.e., bigger image size, higher granularity) than transformer models
- Use cross-attention between text input and diffusional model

Stable Diffusion

'A painting of the last supper by Picasso.'



Stable Diffusion: Rombach et al. (2022), “High-Resolution Image Synthesis with Latent Diffusion Models”

Social NLP

Peter Berglund:

You're still going to have to convince a jury that I killed two strangers for no reason.

Grissom doesn't look worried.

He takes his gloves off and puts them on the table.

Grissom:

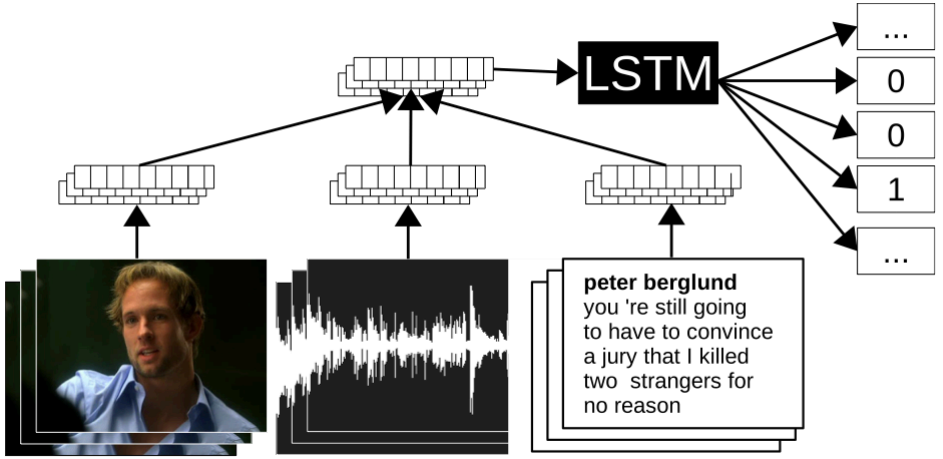
You ever been to the theater Peter?

There 's a play called six degrees of separation.

Frermann et al. 2018, “Whodunnit? Crime Drama as a Case for Natural Language Understanding”

Social NLP

- LSTM-based model
- Video/audio/script input as features at each time step
- Output: is the perpetrator mentioned in that time step?



Frermann et al. 2018, “Whodunnit? Crime Drama as a Case for Natural Language Understanding”

Week	Date	Topic	Readings
1	1/17	Introduction [slides]	Browse http://nlpprogress.com
	1/18	Lexical semantics/static word embeddings [slides]	SLP3 ch 6
2	1/24	Text classification 1: Logistic regression [slides]	SLP3 ch 5
	1/26	Text classification 2: MLP and CNN [slides]	SLP3 ch 7.1-7.4; 7.6; G ch. 13
3	1/31	Text classification 3: Attention and transformers [slides]	SLP3 ch 10
	2/2	Annotation [slides]	PS ch. 6
4	2/7	Language modeling 1 [slides]	SLP3 ch 3; SLP3 ch 7.5; 7.7; SLP ch 9.2
	2/9	Language modeling 2: Contextual embeddings [slides]	SLP3 ch 10
5	2/15	Language modeling 3: Prompting methods and reinforcement learning from human feedback (RLHF) [slides]	Liu et al. 2021; Ouyang et al. 2022
	2/16	Sequence labeling: POS tagging; HMM [slides]	SLP3 ch 8
6	2/21	Neural sequence labeling [slides]	SLP3 ch 9.3-9.6
	2/23	Midterm 1 (in class)	
7	2/28	Syntax [slides]	SLP3 ch 17; SLP3 ch 18

DB

Write an episode of Game of Thrones set in UC Berkeley.

Title: "The Battle for Berkeley"

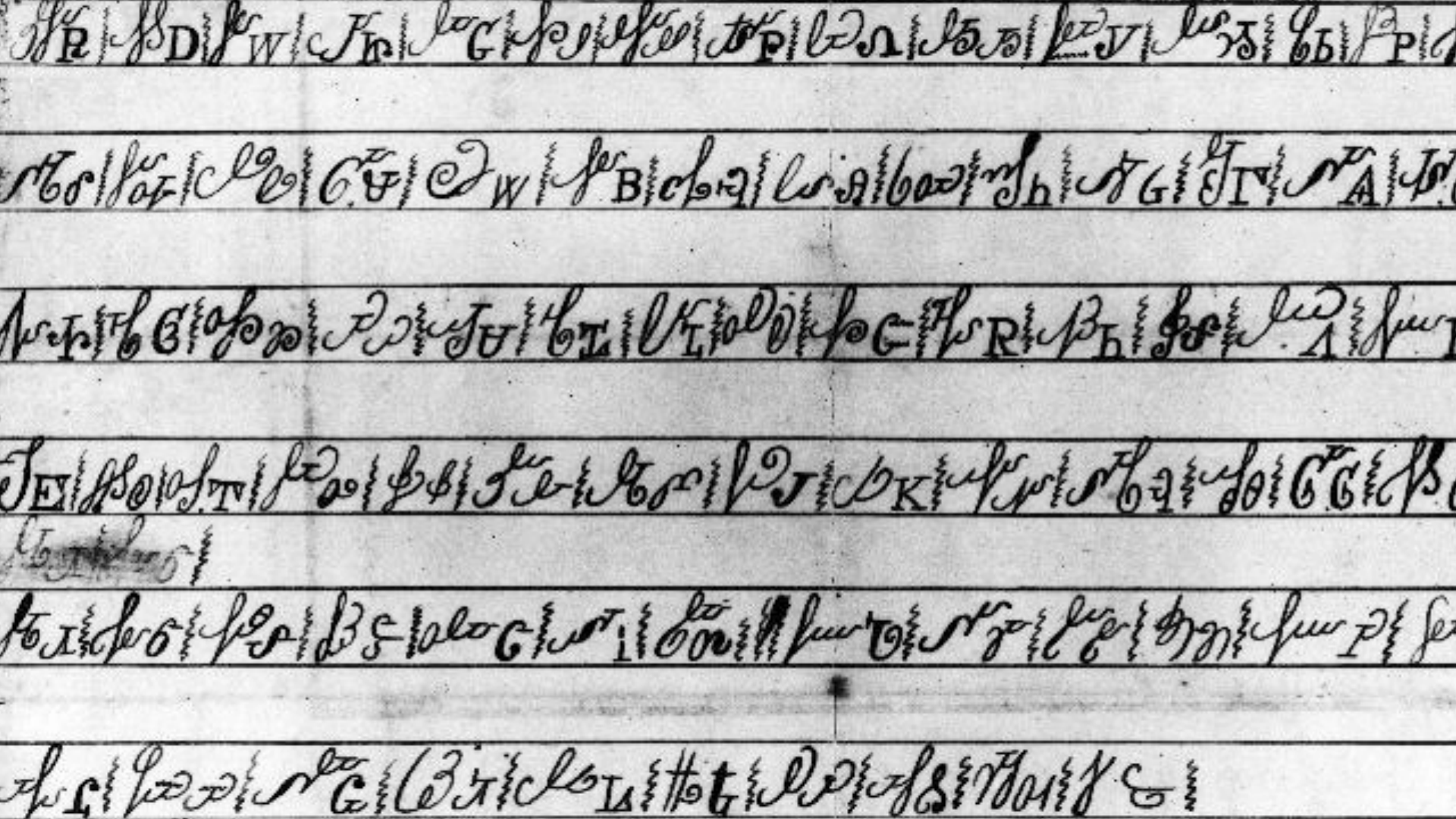
INT. UC BERKELEY CAMPUS - DAY

The campus of UC Berkeley is bustling with students going about their daily activities. However, tension is in the air as different student groups, representing various factions, gather for a protest in Sproul Plaza. Among them are Sansa Stark, a fierce student activist, and Tyrion Lannister, a witty and cunning political science professor.

SANSA
(leading the protest)
The administration cannot silence our voices! We demand justice and equality for all students!

TYRION
(smiling)
Well said, my lady. Your passion and determination are truly admirable.


“an armchair in the shape of an avocado.”
https://openai.com/blog/dall-e/




OAKLAND POLICE
1770
5005

Questions!?

Course evaluations!

https://course-evaluations.berkeley.edu/berkeley/