Image Captioning

A group of people shopping at an outdoor market.

There are many vegetables at the fruit stand.

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?

Visual QA

“How many horses are in this image?”

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...

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...

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) a tapir with the texture of an accordion.
<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Topic</th>
<th>[slides]</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/18</td>
<td>Introduction</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1/20</td>
<td>Lexical semantics/static word embeddings</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1/25</td>
<td>Text classification 1: Logistic regression</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1/27</td>
<td>Text classification 2: MLP and CNN</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2/1</td>
<td>Text classification 3: Attention and transformers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2/3</td>
<td>Annotation</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2/8</td>
<td>Language modeling 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2/10</td>
<td>Language modeling 2</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2/15</td>
<td>Language modeling 3: Contextual embeddings</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2/17</td>
<td>Language modeling 4: Few-shot learning and prompting methods</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2/22</td>
<td>Sequence labeling: POS tagging; HMM</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2/24</td>
<td>Neural sequence labeling</td>
<td></td>
</tr>
</tbody>
</table>
Text is data.
“an armchair in the shape of an avocado.”
https://openai.com/blog/dall-e/
The importance of being on twitter

by Jerome K. Jerome
London, Summer 1897

It is a curious fact that the last remaining form of social life in which the people of London are still interested is Twitter. I was struck with this curious fact when I went on one of my periodical holidays to the sea-side, and found the whole place twittering like a starling-cage. I called it an anomaly, and it is.

I spoke to the sexton, whose cottage, like all sexton’s cottages, is full of antiquities and interesting relics of former centuries. I said to him, "My dear sexton, what does all this twittering mean?" And he replied, "Why, sir, of course it means Twitter."
"Ah!" I said, "I know about that. But what is Twitter?"

"It is a system of short and pithy sentences strung together in groups, for the purpose of conveying useful information to the initiated, and entertainment and the exercise of wits to the initiated, and entertainment and the exercise of wits to the rest of us."
Course evaluations!

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