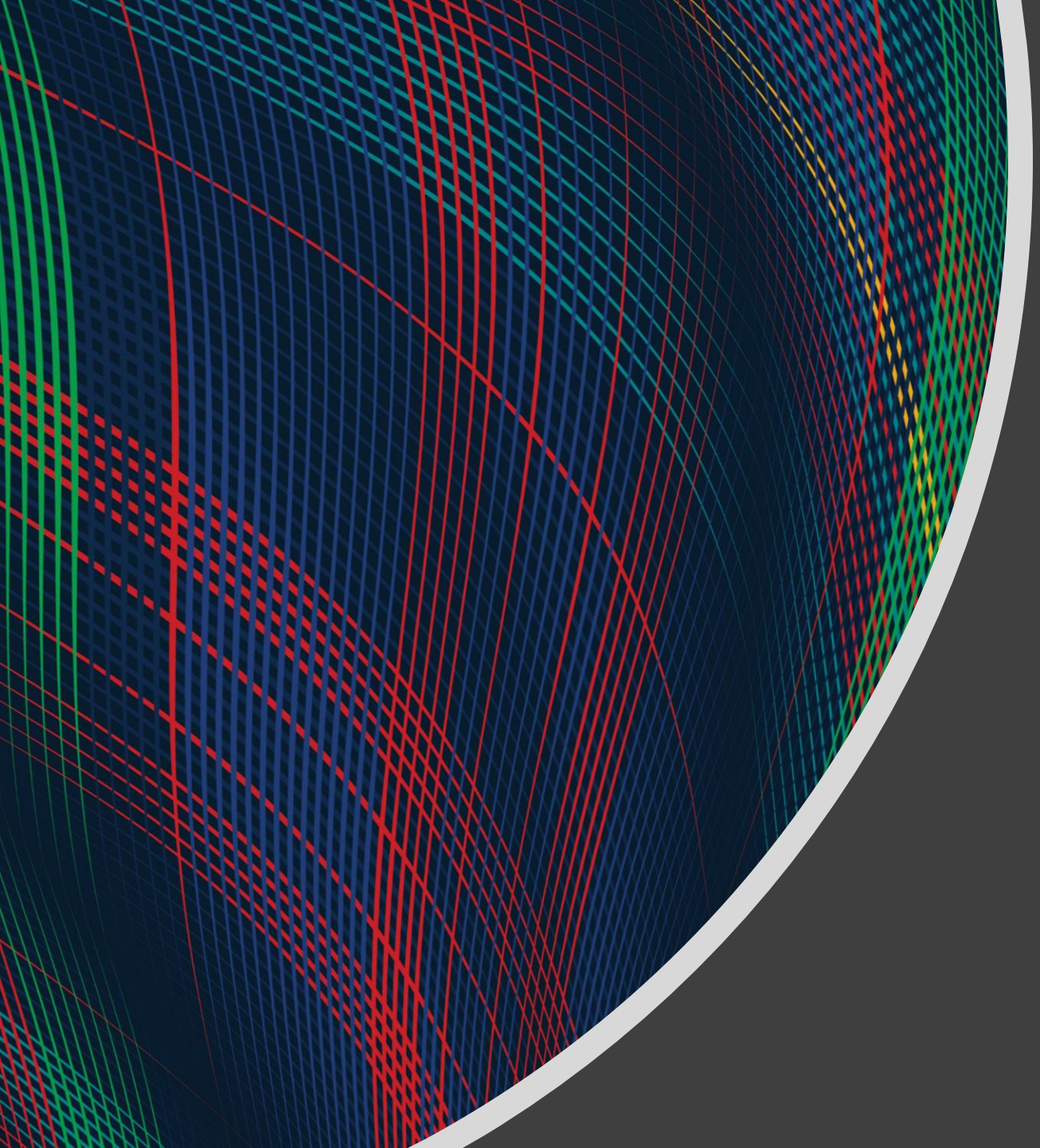


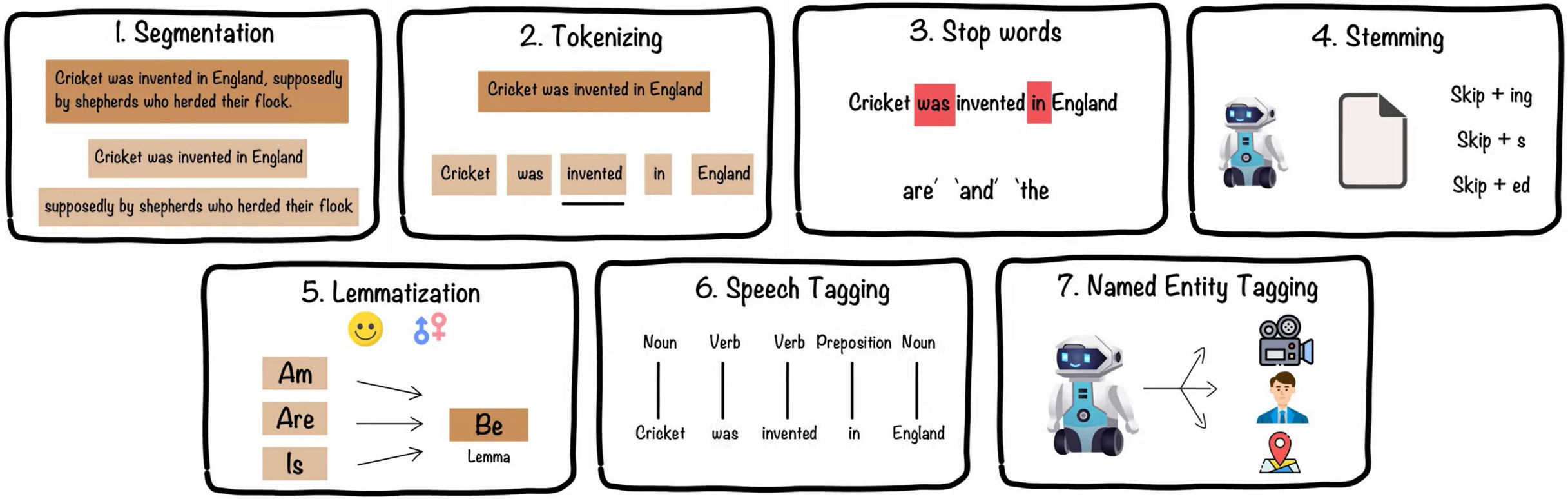
Demystifying AI

Natural Language Processing

Instructor: Pat Virtue

NLP Sub-tasks

Many different operations to help process language



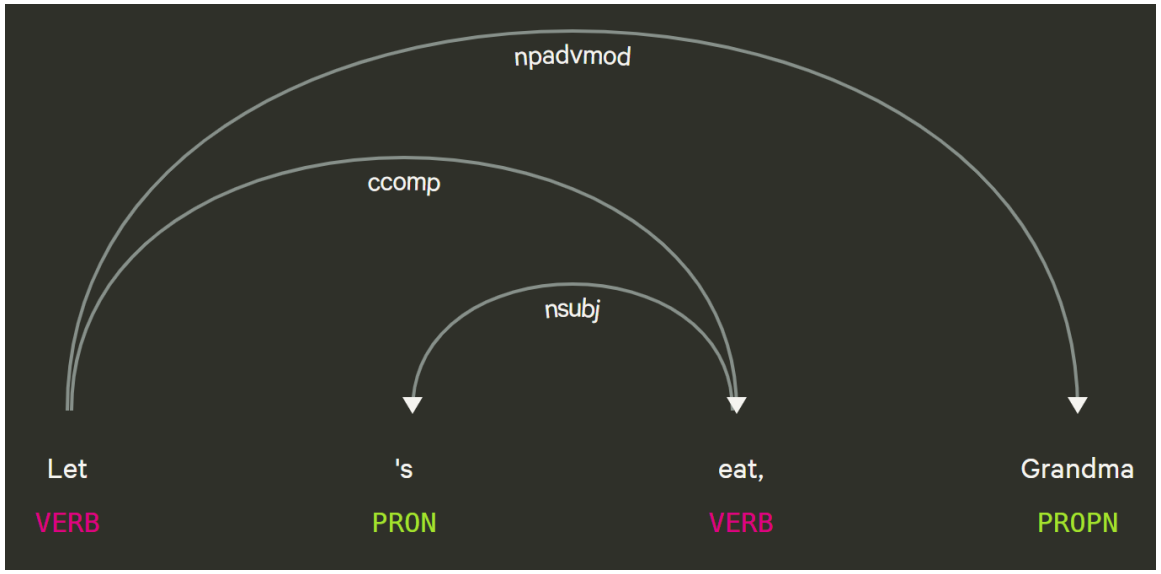
Images: https://www.youtube.com/watch?v=CMrHM8a3hqw

NLP Grammar

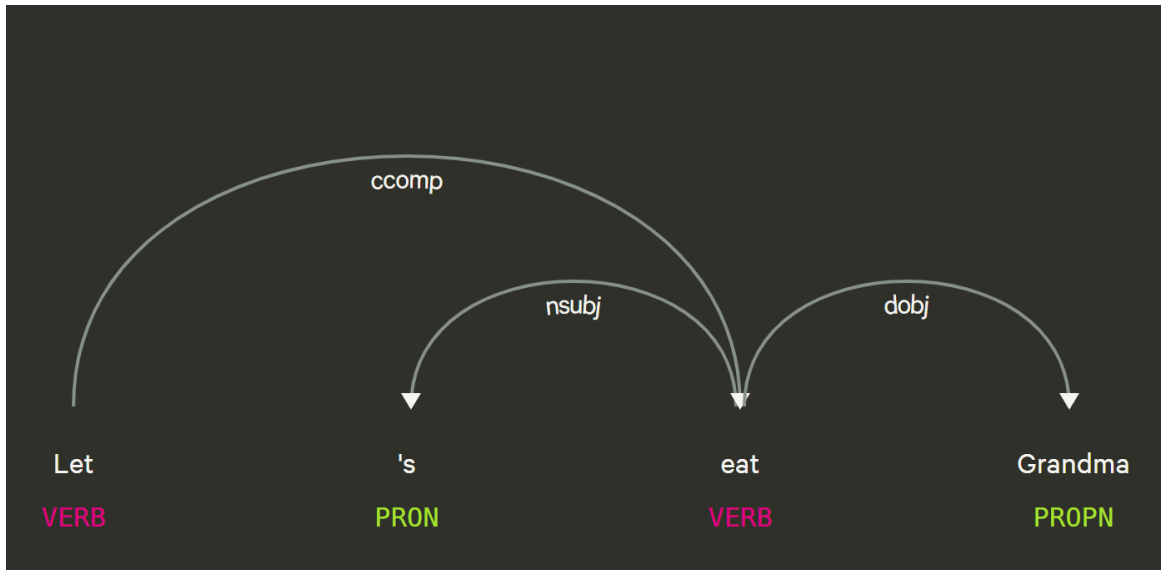
Text to parts of speech and parse tree

https://explosion.ai/demos/displacy

Let’s eat, Grandma



Let’s eat Grandma



Commas save lives ☺

Language is Hard

Emphasis can drastically change meaning

I didn’t eat your dog

Exercise: NLP Tasks

How many different NLP Input/Output agents can you think of?

Sentiment Analysis

Sentiment analysis demo

https://text2data.com/Demo

“I recommend that you find something better to do with your time”

NLP Features

Hand-crafted features

NLP Features

Word → integer index in vocabulary list

NLP Features

Word embeddings

Training data:

“The king sat on the throne”

“the queen sat on the throne”

“the banana is yellow”

“they sat on the yellow bus”

- king
- sat
- throne
- queen
- banana
- yellow
- they
- bus

- king
- sat
- throne
- queen
- banana
- yellow
- they
- bus

Skip-gram

```
score(word, <other words around it>)
```

NLP Features

Word embeddings

Training data:

“The king sat on the throne”

“the queen sat on the throne”

“the banana is yellow”

“they sat on the yellow bus”

- king
- sat
- throne
- queen
- banana
- yellow
- they
- bus

- king
- sat
- throne
- queen
- banana
- yellow
- they
- bus

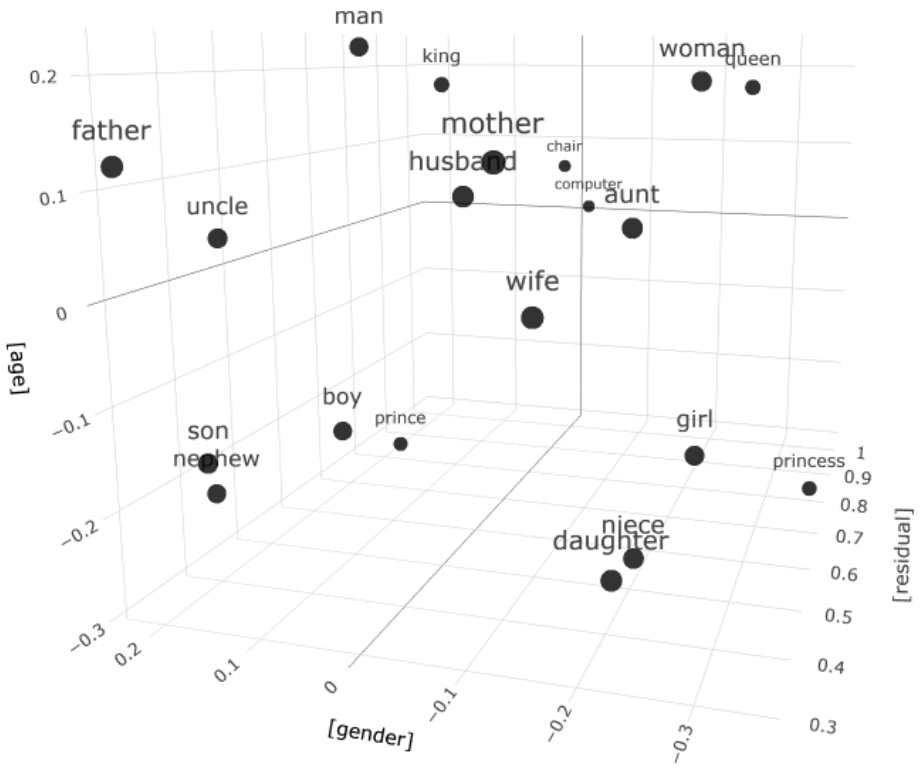
Skip-gram

```
score(word, <other words around it>)
```

NLP Features

Word embeddings demo

https://www.cs.cmu.edu/~dst/test/Word2VecDemo/index.html



Human NLP

Wheel of Fortune

Probability Models

Example: Speech Recognition

“artificial ……”

Find most probable next word given “artificial” and the audio for second word.

Probability Models

Example: Speech Recognition

“artificial ……”

Find most probable next word given “artificial” and the audio for second word.

Which second word gives the highest probability?

$P(\textbf{limb} \mid \text{artificial}, \text{audio})$

$P(\textbf{intelligence} \mid \text{artificial}, \text{audio})$

$P(\textbf{flavoring} \mid \text{artificial}, \text{audio})$

<u>Break down problem</u>

n-gram probability * audio probability

$P(\textbf{limb} \mid \text{artificial}) * P(\text{audio} \mid \textbf{limb})$

$P(\textbf{intelligence} \mid \text{artificial}) * P(\text{audio} \mid \textbf{intelligence})$

$P(\textbf{flavoring} \mid \text{artificial}) * P(\text{audio} \mid \textbf{flavoring})$

N-gram Training

Where do the n-gram probabilities come from?

Google n-grams demo

NLP Can Be Huge

N-gram probabilities

Vocabulary size: 50,000

NLP Training

Self-supervised

Example: Jane Austen, *Pride and Prejudice*

Vanity and pride are different things, though the words are often used synonymously. A person may be proud without being vain. Pride relates more to our opinion of ourselves, vanity to what we would have others think of us.

NLP Training

Self-supervised learning (auto-regressive)

Example: Jane Austen, *Pride and Prejudice*

Vanity and pride are different things, though the words are often used synonymously. A person may be proud without being vain. Pride relates more to our opinion of ourselves, vanity to what we would have others think of us.

Examples

Random samples from language model trained on Shakespeare:

n=1: "in as , stands gods revenge ! france pitch good in fair hoist an what fair shallow-rooted , . that with wherefore it what a as your . , powers course which thee dalliance all"

n=2: "look you may i have given them to the dank here to the jaws of tune of great difference of ladies . o that did contemn what of ear is shorter time ; yet seems to"

n=3: "believe , they all confess that you withhold his levied host , having brought the fatal bowels of the pope ! ' and that this distemper'd messenger of heaven , since thou deniest the gentle desdemona ,"

n=7: "so express'd : but what of that ? 'twere good you do so much for charity . i cannot find it ; 'tis not in the bond . you , merchant , have you any thing to say ? but little"

This is starting to look a lot like Shakespeare… because it is Shakespeare

Slide: CMU, Zico Kolter

GPT-3 Language Model

Advanced language model

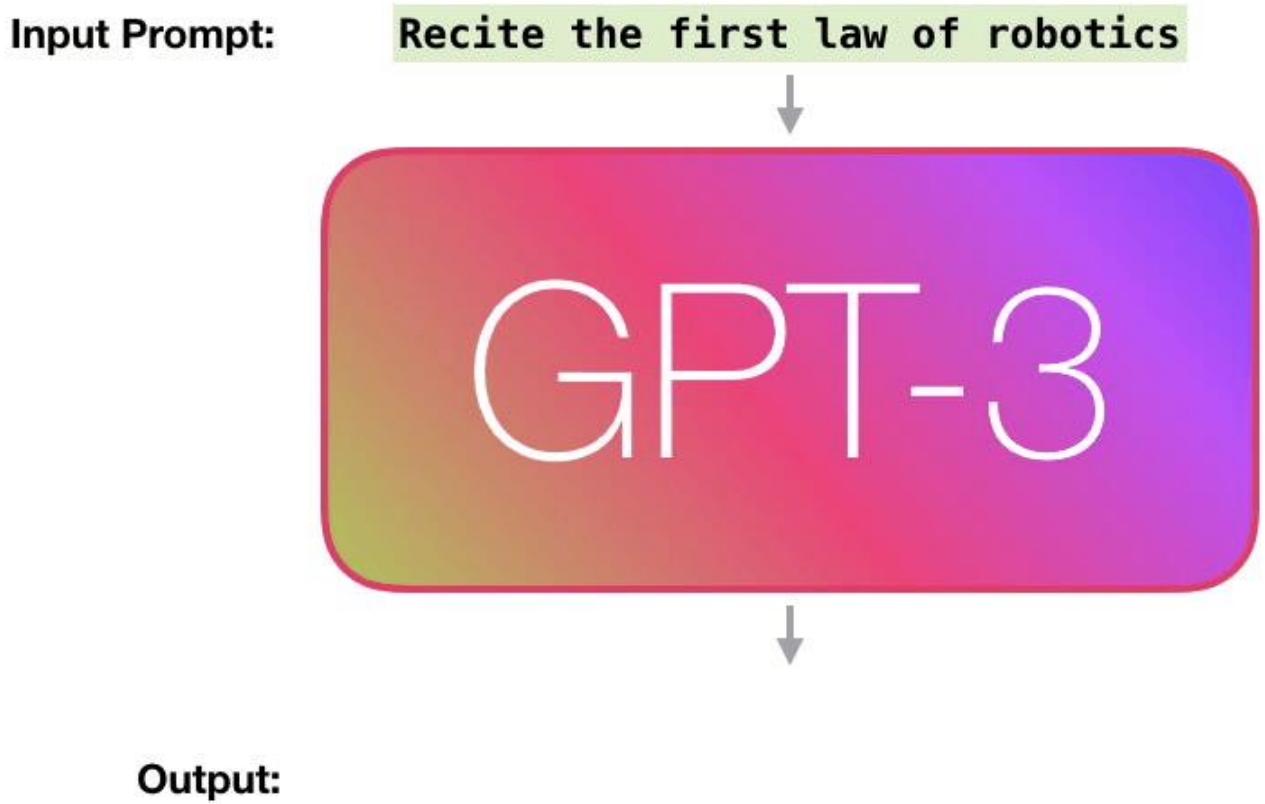
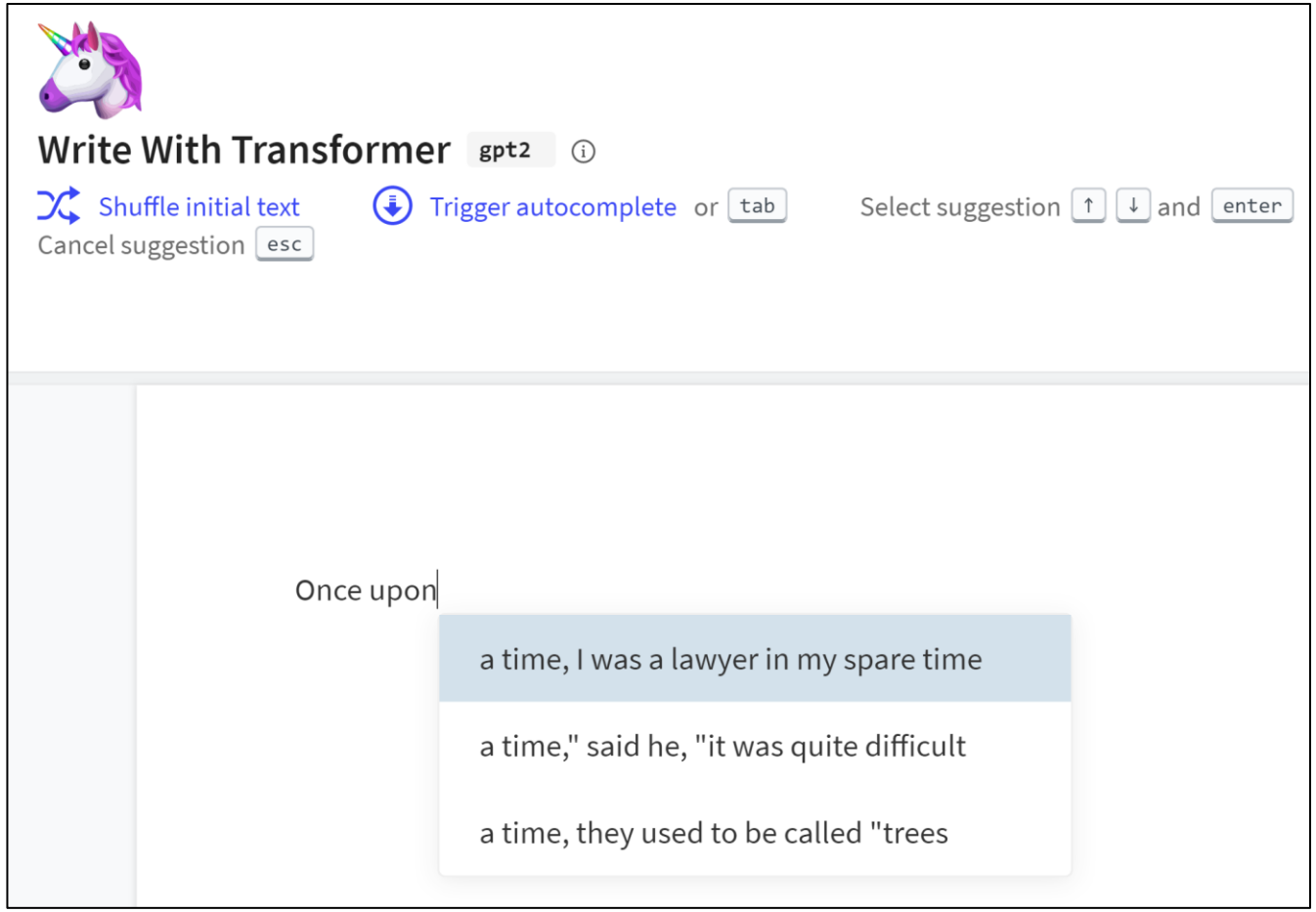


Image: https://jalammar.github.io/how-gpt3-works-visualizations-animations/

Language Model

Generate highly probable next words

https://transformer.huggingface.co/doc/gpt2-large



Language Model with Images

Create new images from text

https://openai.com/blog/dall-e/

https://openai.com/dall-e-2/

an armchair in the shape of an avocado. . . .

GPT-3 Language Model

State-of-the-art language model

- Trained with dataset of 300 billion tokens
- 175 billion parameters
- It was estimated to cost 355 GPU years and cost $4.6m

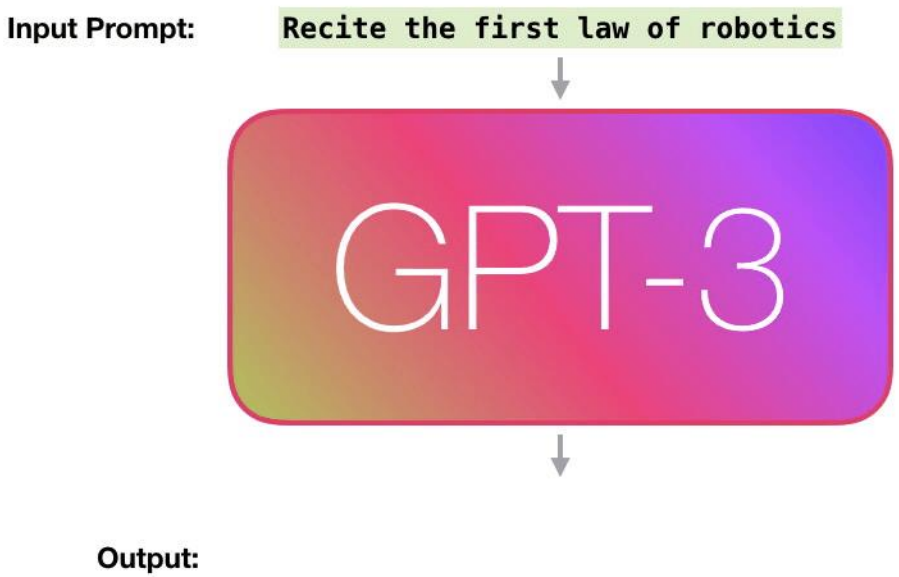
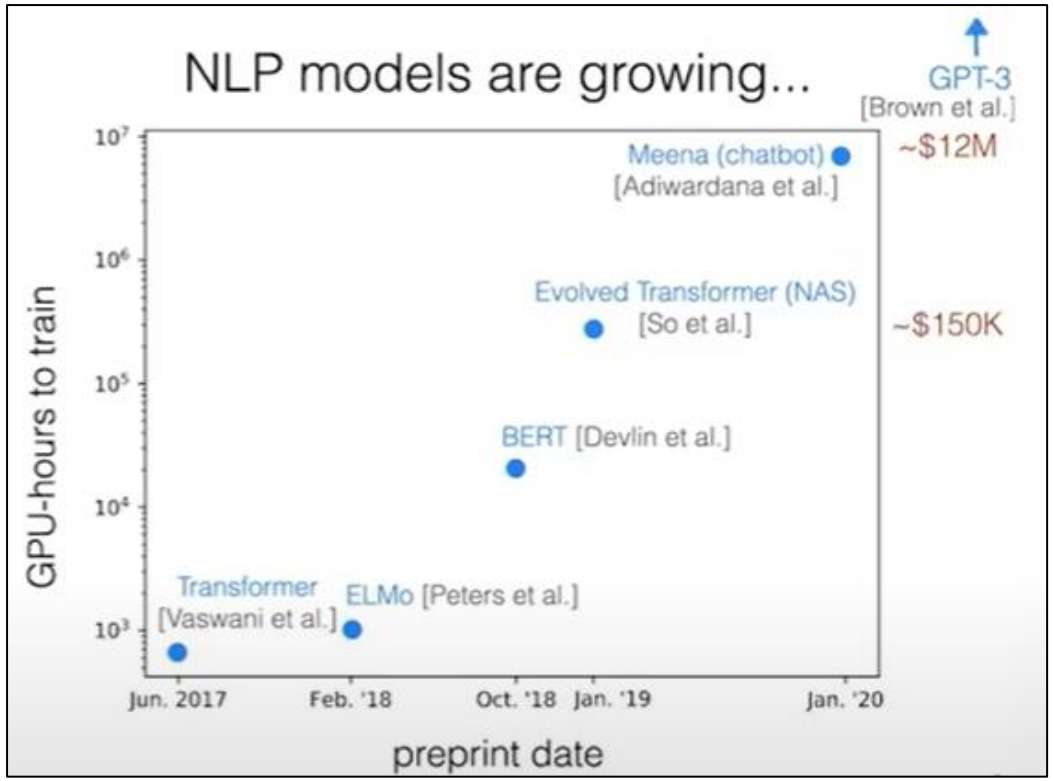


Image: https://jalammar.github.io/how-gpt3-works-visualizations-animations/

NLP Ethics

Environmental concerns

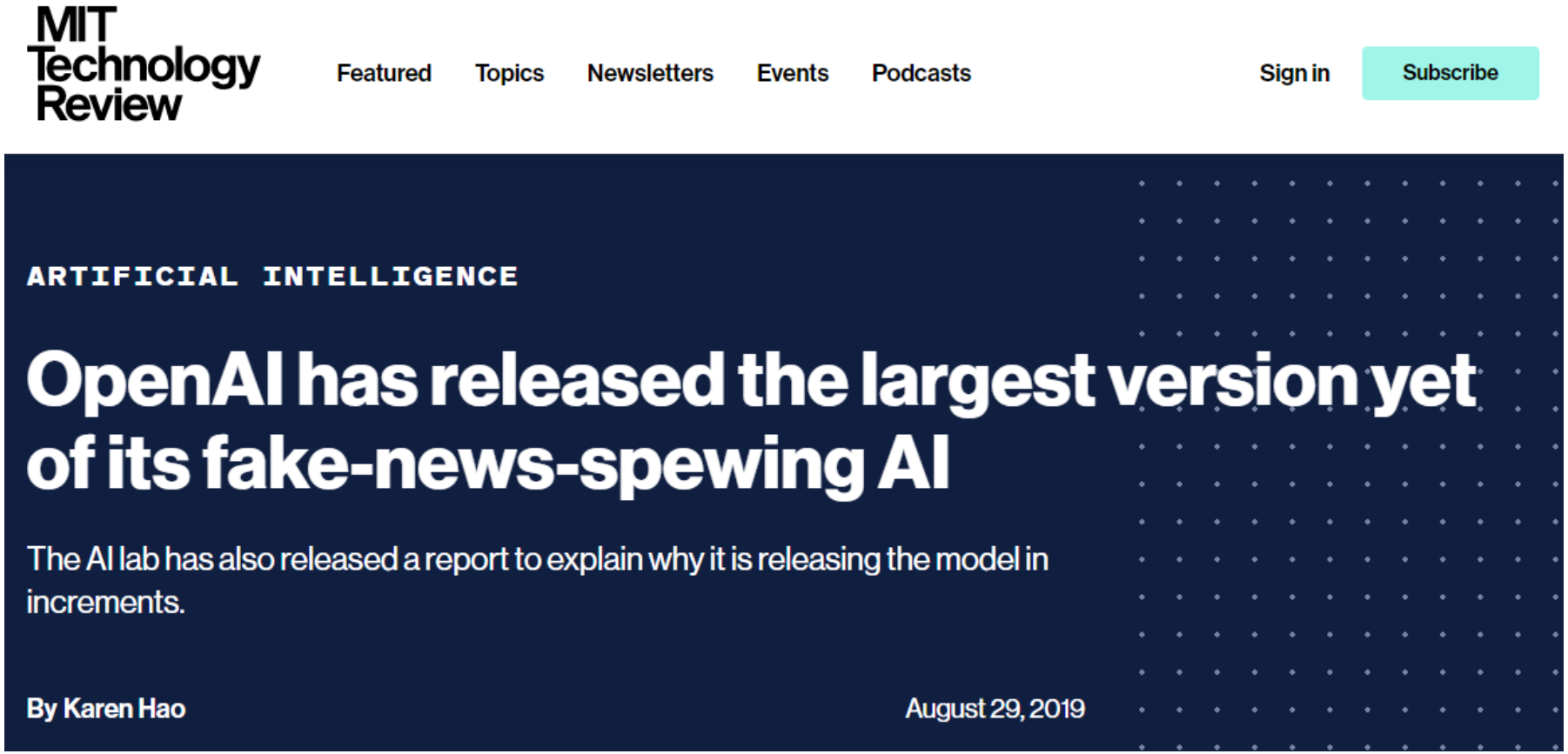


Emma Strubell
Assistant Professor
Language Technologies Institute
CMU

[Frontiers in Machine Learning: Climate Impact of Machine Learning]

NLP Ethics

Misuse



https://www.technologyreview.com/2019/08/29/133218/openai-released-its-fake-news-ai-gpt-2/

NLP Ethics

Bias

https://spectrum.ieee.org/tech-talk/artificial-intelligence/machine-learning/in-2016-microsofts-racist-chatbot-revealed-the-dangers-of-online-conversation

NLP Ethics

Bias

https://venturebeat.com/2021/06/10/openai-claims-to-have-mitigated-bias-and-toxicity-in-gpt-3/

NLP Ethics

Dangerous errors

AINEWS

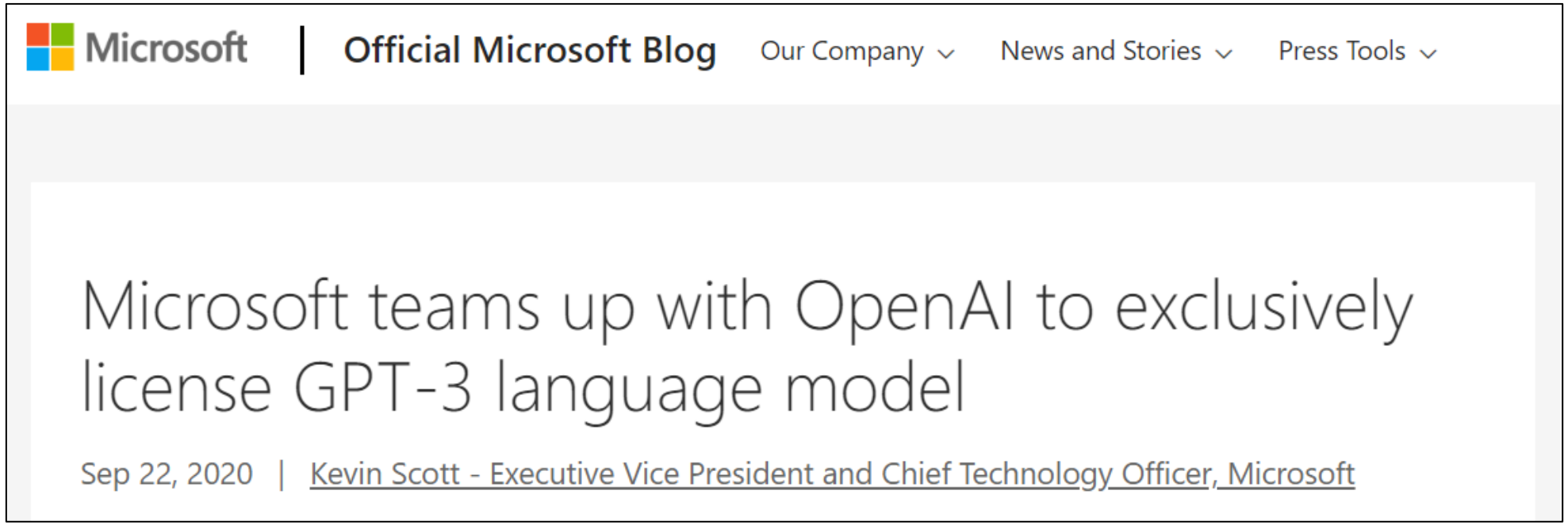
Medical chatbot using OpenAI's GPT-3 told a fake patient to kill themselves

By Ryan Daws | October 28, 2020 | TechForge Media
Categories: Chatbots, Healthcare,

https://artificialintelligence-news.com/2020/10/28/medical-chatbot-openai-gpt3-patient-kill-themselves/

NLP Ethics

Digital divide



https://blogs.microsoft.com/blog/2020/09/22/microsoft-teams-up-with-openai-to-exclusively-license-gpt-3-language-model/

NLP Ethics

Discussion

Should a company be required to share their powerful, trained AI model (GPT-3) with the public?

- Pros:

- Cons:

- Conclusion?