Computer Vision: Fall 2022 — Lecture 18
Dr. Karthik Mohan

Univ. of Washington, Seattle

December 6, 2022

References

Generic ML/DL

1. Good Book for Machine Learning Concepts
2. Deep Learning Reference

CNN

1. Convolutional Neural Networks for Visual Recognition
2. Convolutional Neural Net Tutorial
3. CNN Transfer Learning
4. PyTorch Transfer Learning Tutorial

CNN Publication References

CNN surveys

1. Convolutional Neural Networks: A comprehensive survey, 2019
2. A survey of Convolutional Neural Networks: Analysis, Applications, and Prospects, 2021

CNN Archs

1. GoogLeNet
2. Top models on ImageNet
3. ResNet ILSVRC paper

Object Detection and Image Captioning References

Object Detection

1. A survey of modern deep learning based object detection methods
2. YOLO Survey
3. YOLO Original Paper

Image Captioning

1. From Show to Tell: A survey on Deep Learning-based Image Captioning
2. A survey of image captioning models
3. StyleNet

Generative AI references

Generative Text Models

1. GPT-3 Paper
2. GPT-3 Playground
3. ChatGPT Playground

Today

1. Encoder-Decoder Arch Recap

Today

1. Encoder-Decoder Arch Recap
2. Introduction to GPT-3

Today

1. Encoder-Decoder Arch Recap
2. Introduction to GPT-3
3. GPT-3 and ChatGPT Playground Exploration

Encoder-Decoder Model for Image Captioning
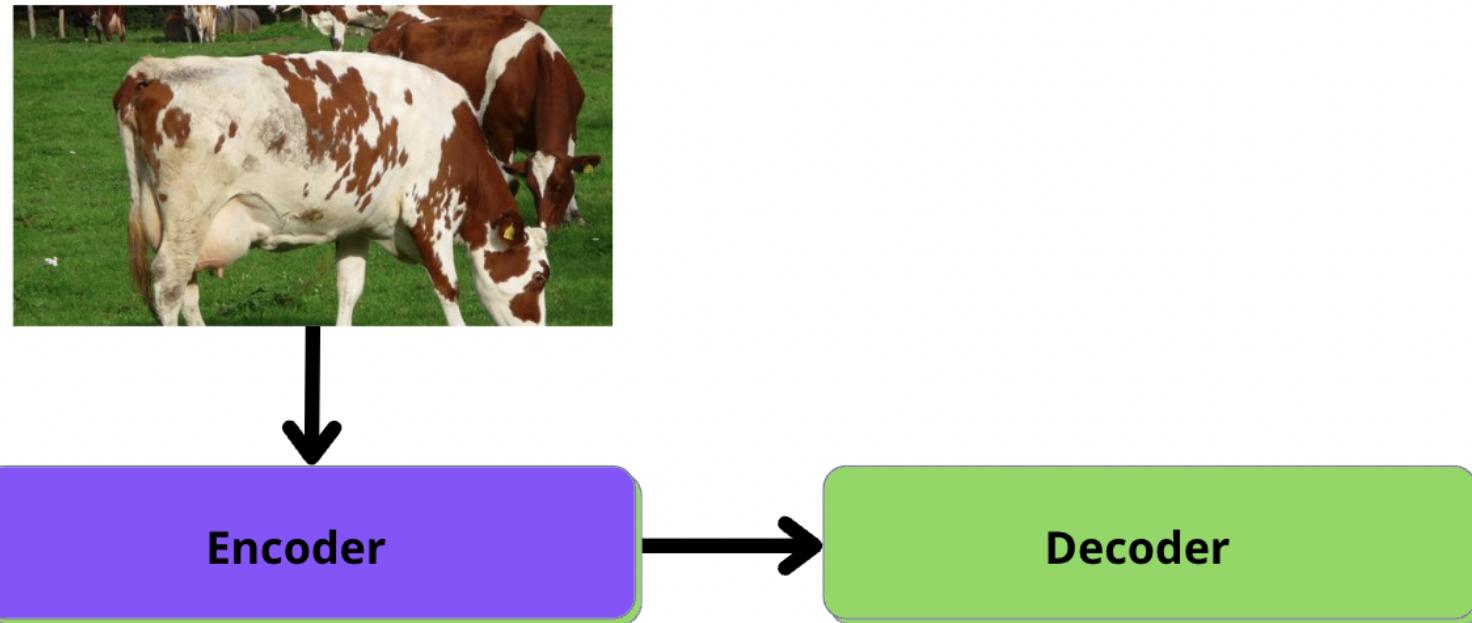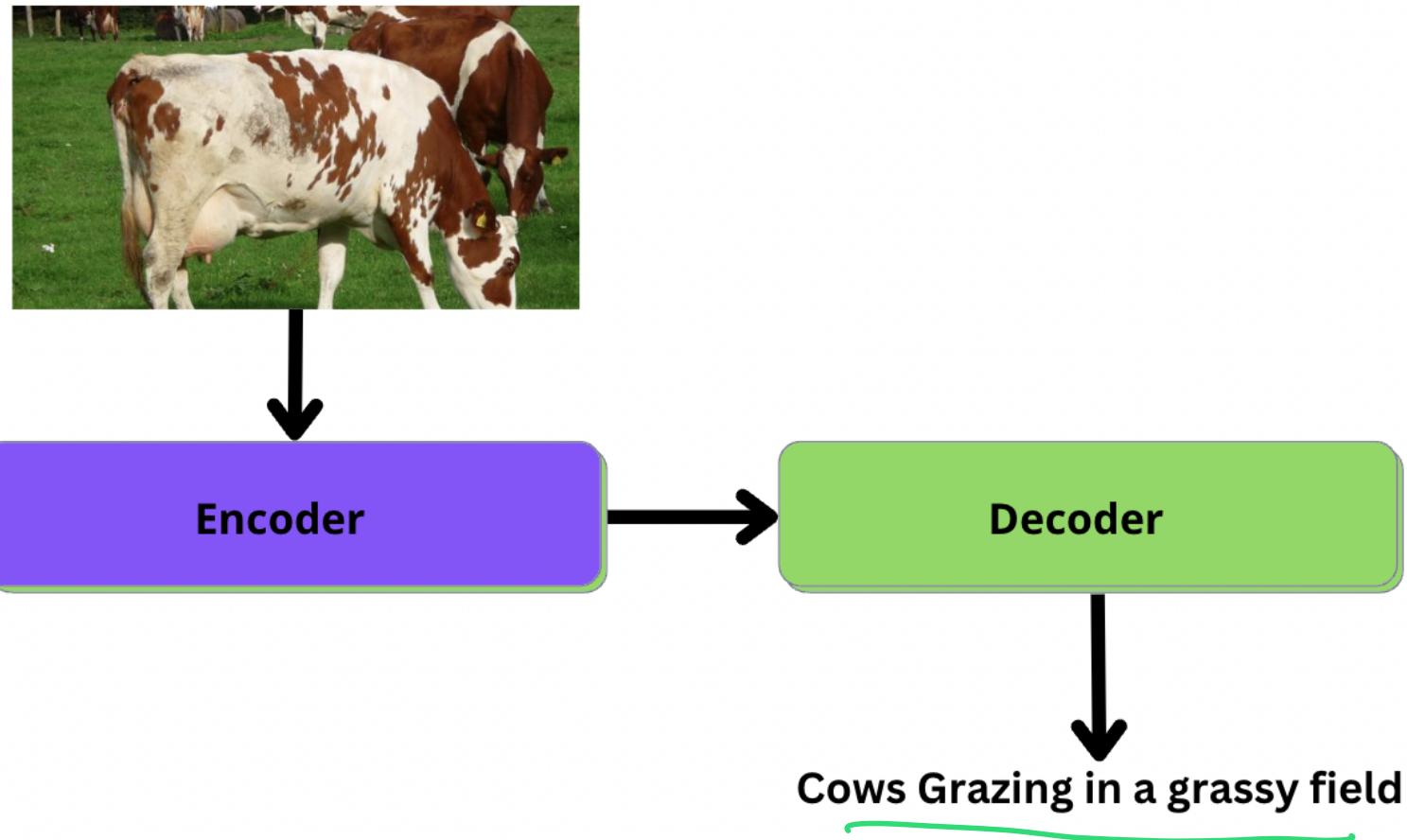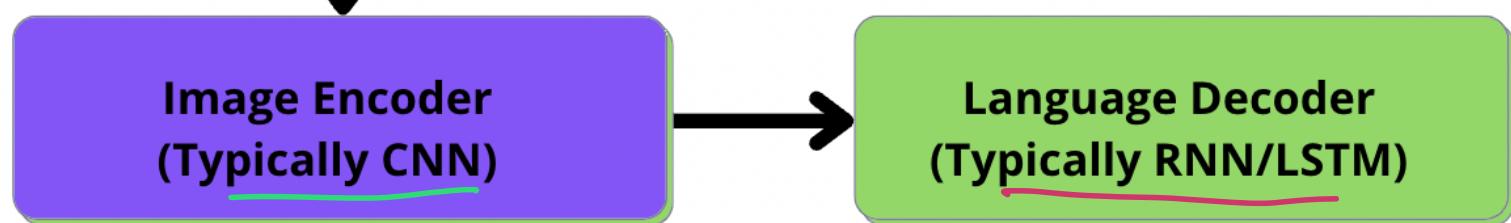
Encoder-Decoder Model for Image Captioning

Encoder-Decoder Model for Image Captioning
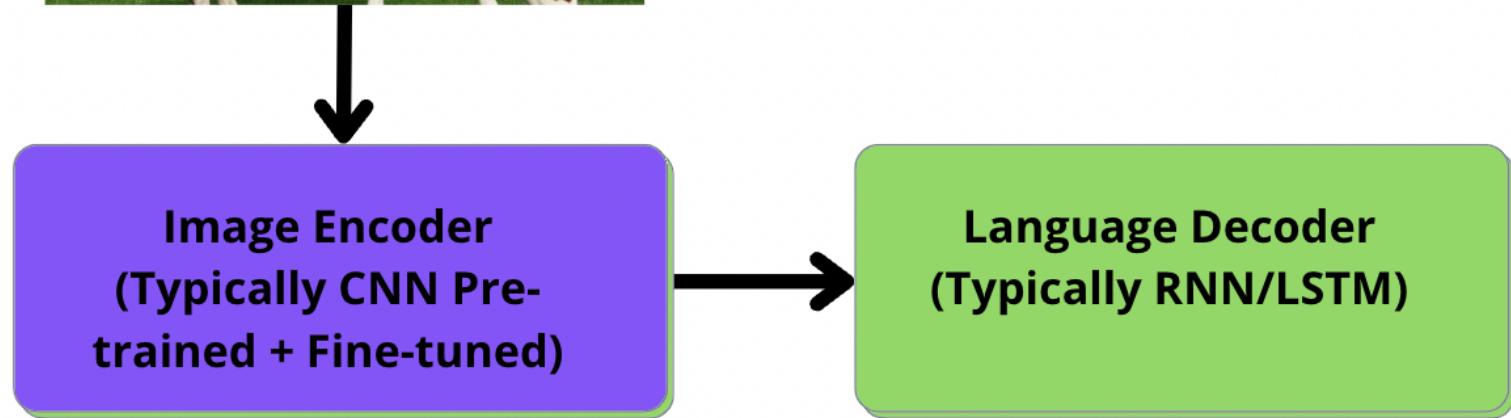
Encoder-Decoder Model for Image Captioning
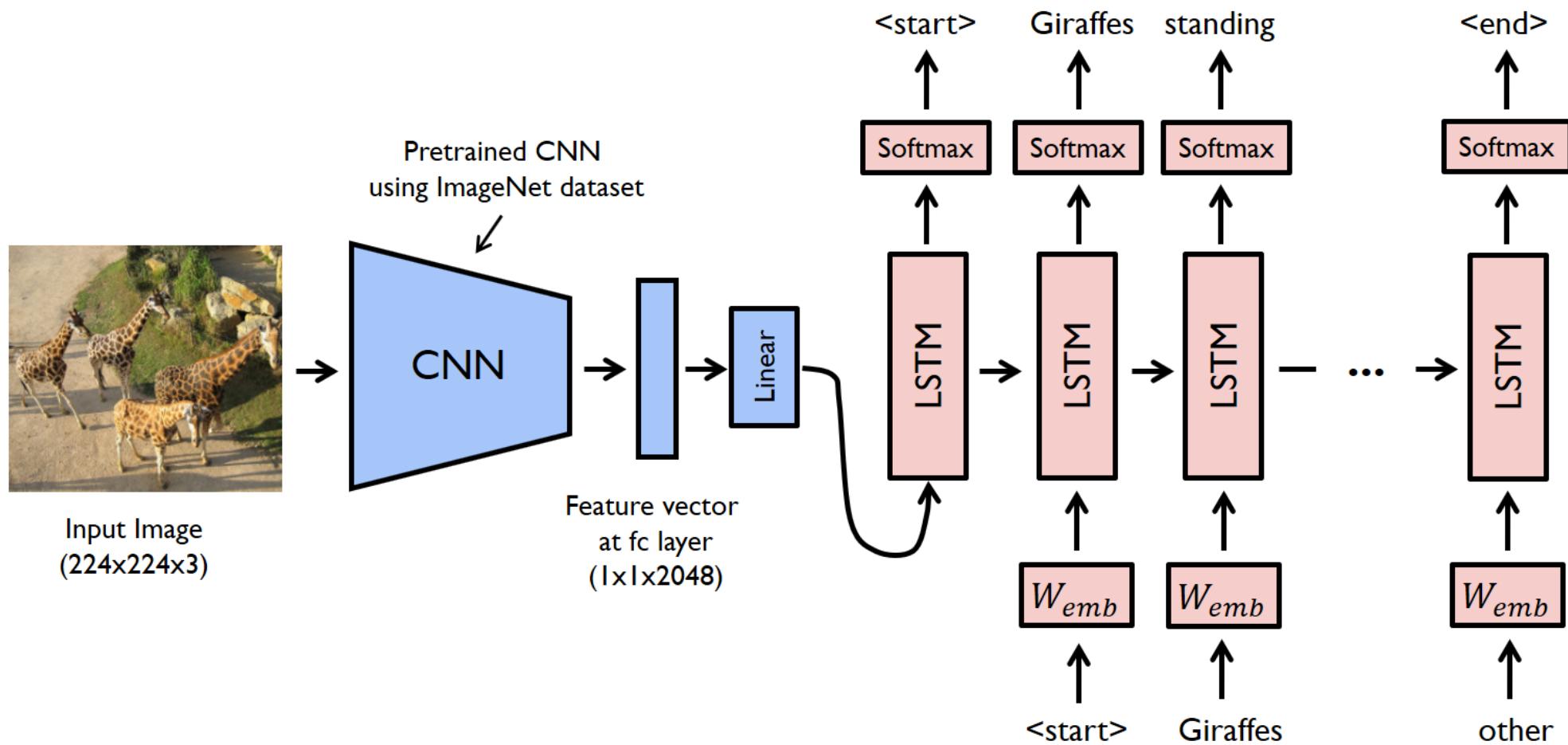
Encoder-Decoder Model for Image Captioning

Encoder-Decoder Model for Image Captioning
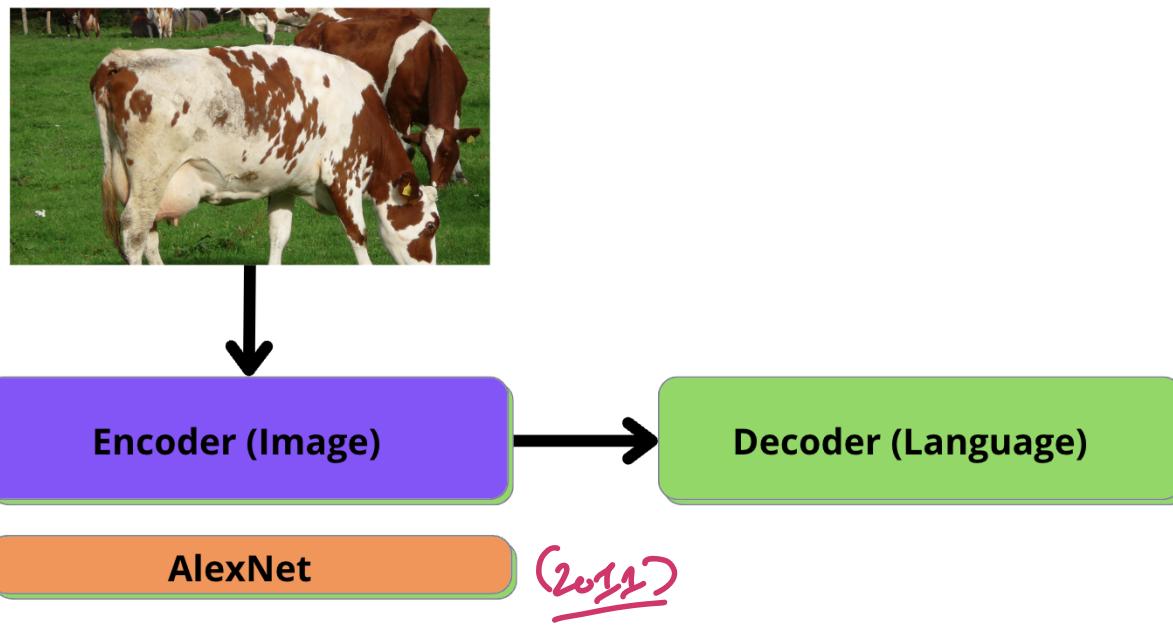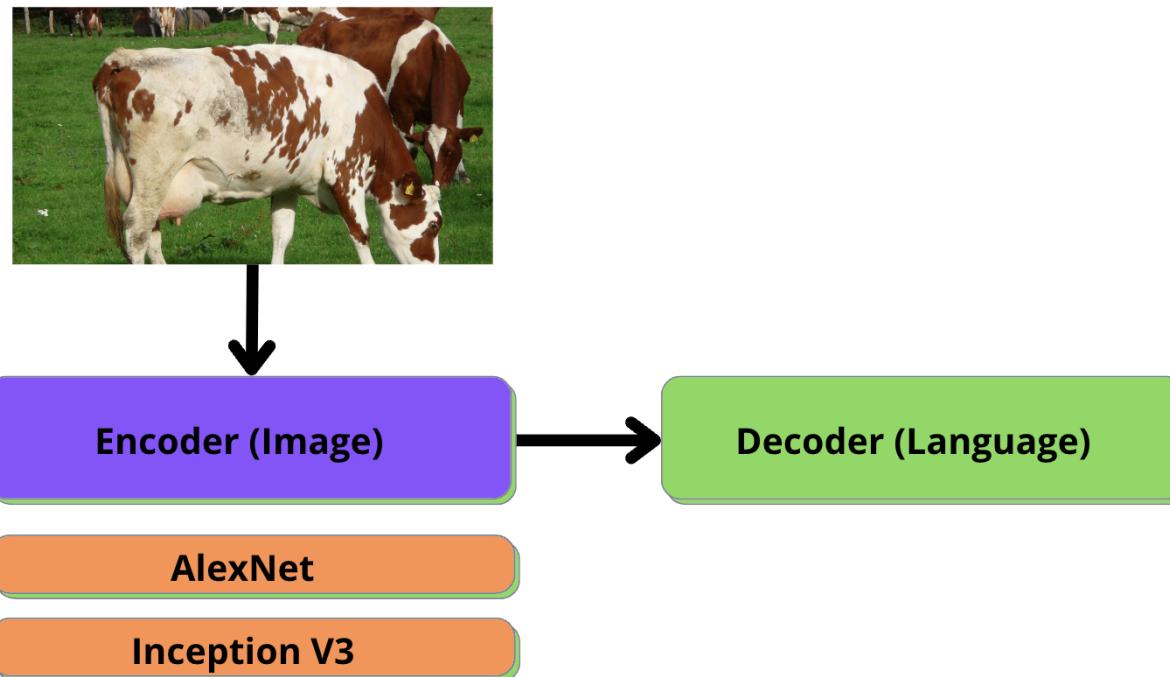
Encoder-Decoder Architecture
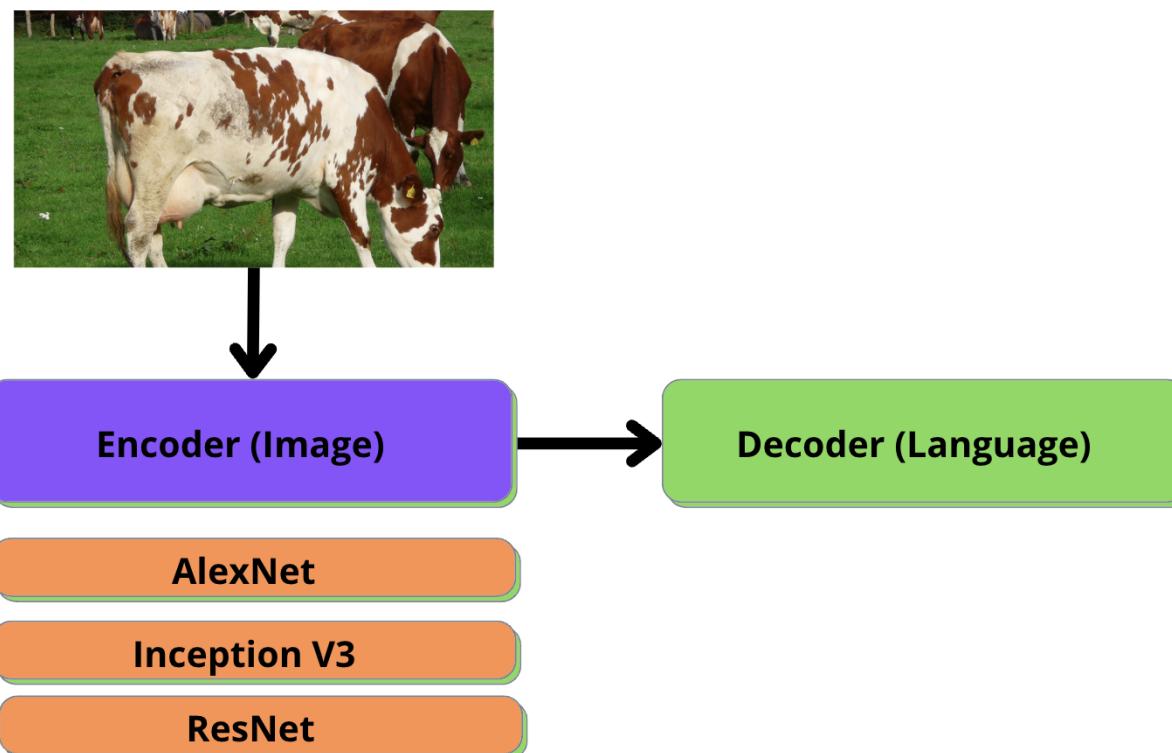
Evolution of Encoders and Decoders
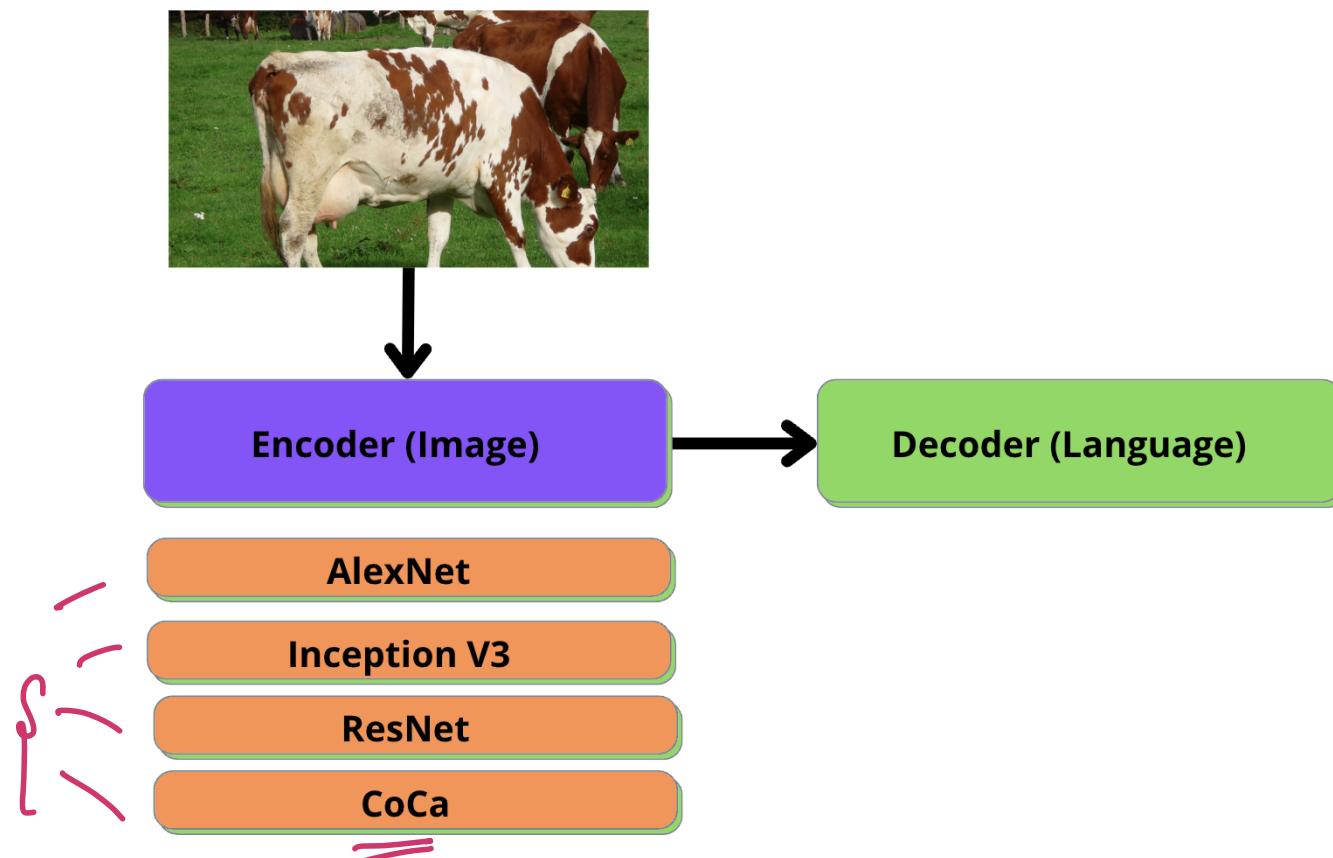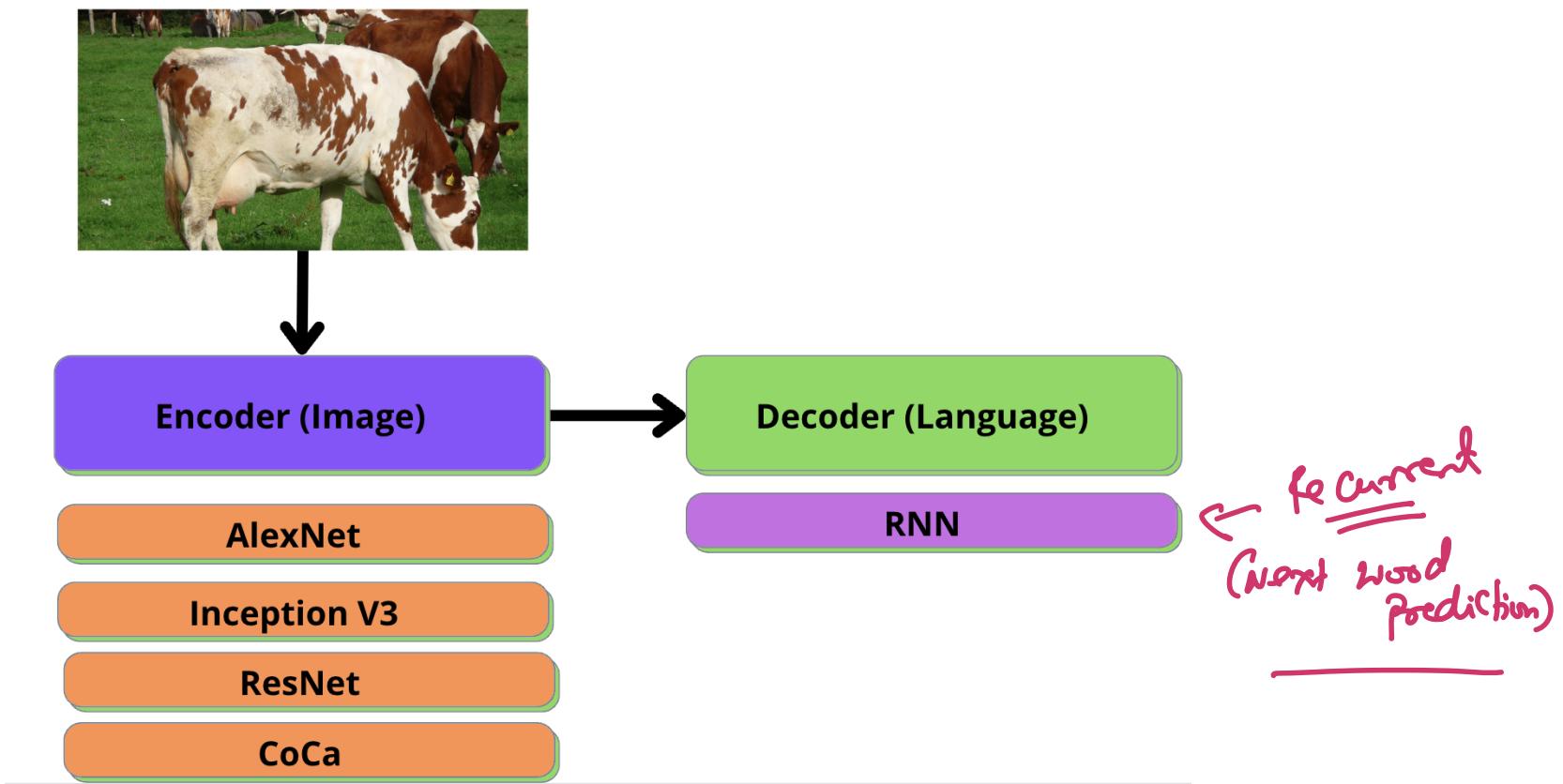
Encoder (Image) → Decoder (Language)

AlexNet (2011)

Evolution of Encoders and Decoders

Evolution of Encoders and Decoders
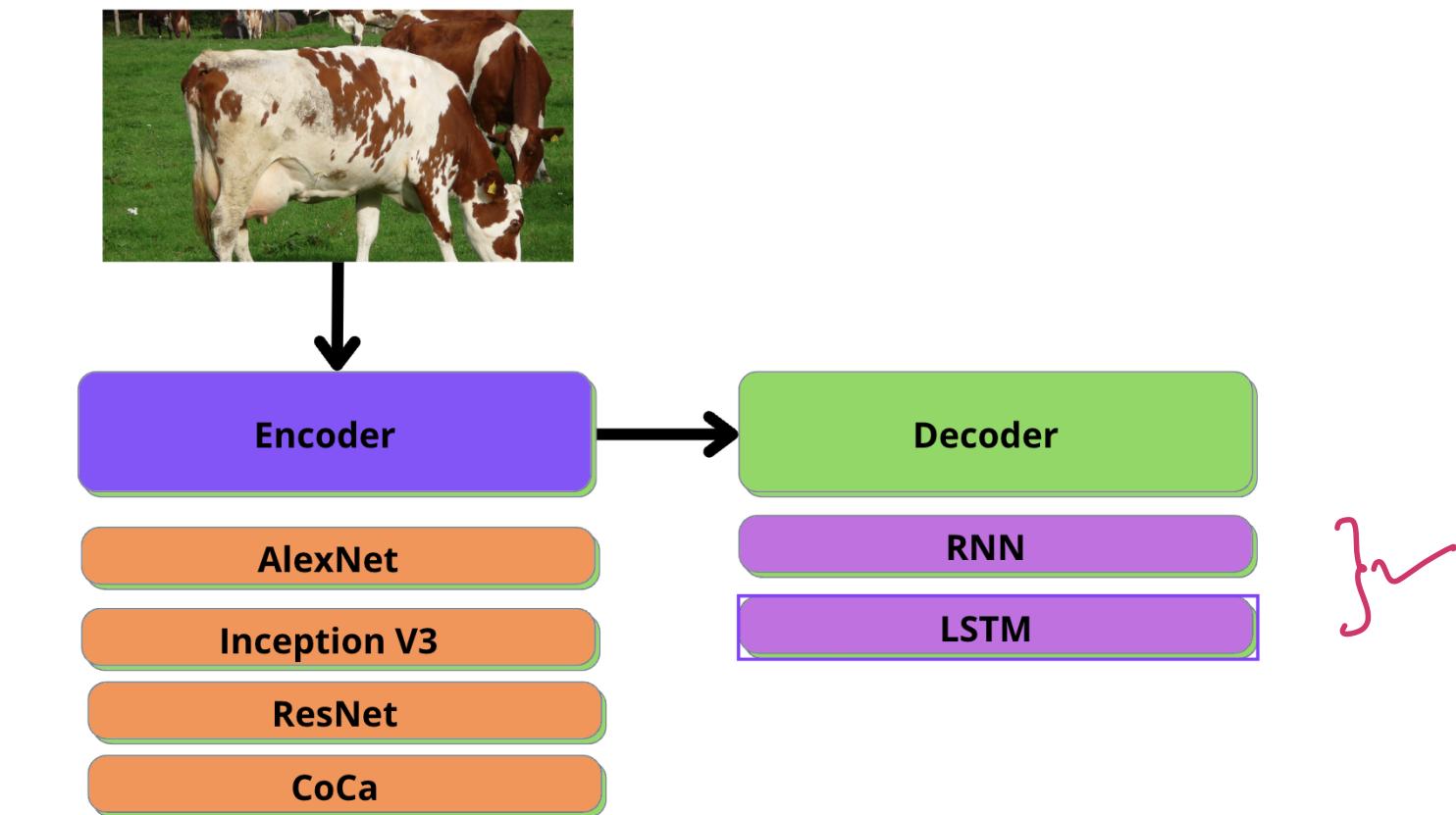
Evolution of Encoders and Decoders
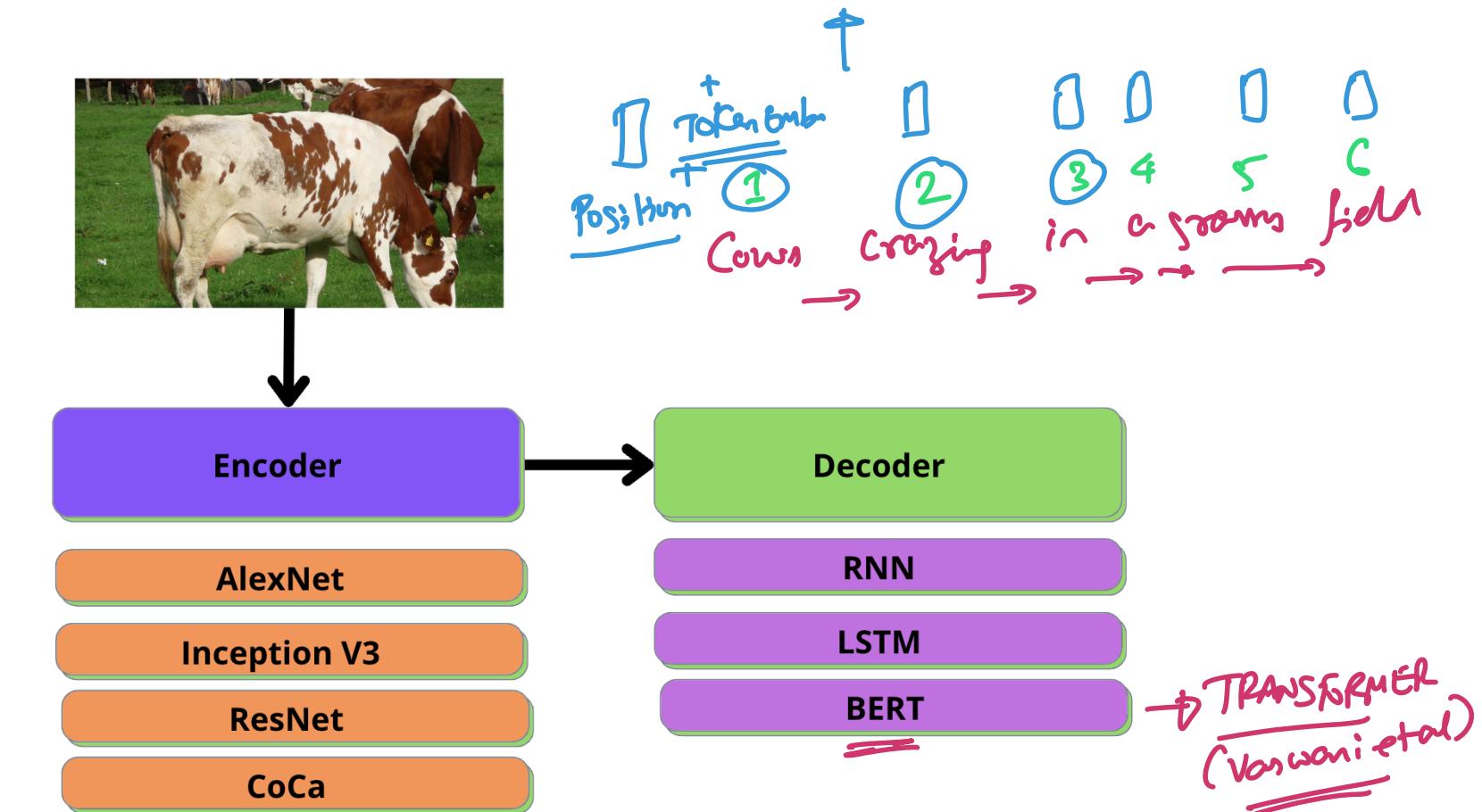
# Evolution of Encoders and Decoders
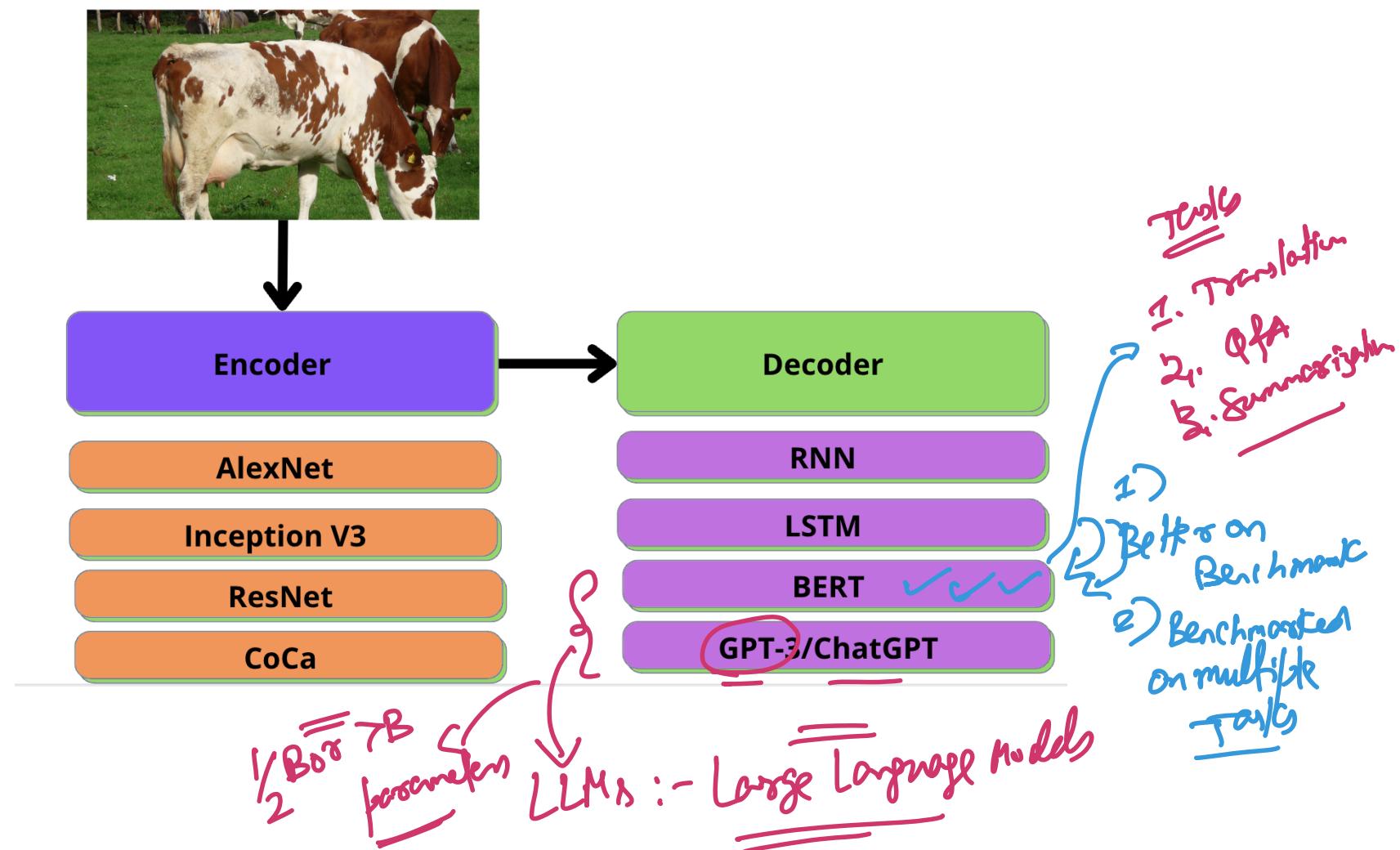
Evolution of Encoders and Decoders

Evolution of Encoders and Decoders

Evolution of Encoders and Decoders

Generative AI from Text - Language and Vision

1. The input is a text description, and the output is a structured document (e.g. code or paragraph or document) - GPT-3 and ChatGPT models are SOTA
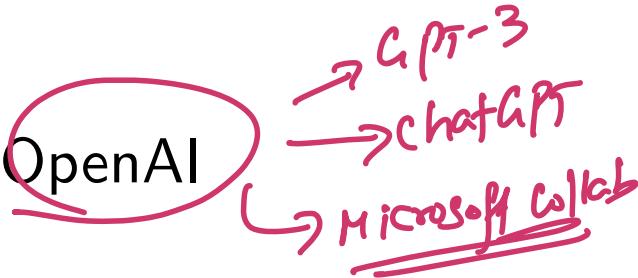
↳ State of the Art

Generative AI from Text - Language and Vision

1. The input is a text description, and the output is a structured document (e.g. code or paragraph or document) - GPT-3 and ChatGPT models are SOTA

2. The input is a text description, and the output is a image - Diffusion Models are SOTA

Next lecture

GPT-3

1. Released in 2020 by OpenAI

GPT-3
→ chatGPT
→ Microsoft collab

GPT-3

① Released in 2020 by OpenAI →BERT

② **Motivation:** Fine-tuning LLMs for specific tasks requires lots of data and generalization issues.

SAT analogy task

man : king → 1 shot learning!

women : ? queen

↑ test example

GPT-3

1. Released in 2020 by OpenAI
2. **Motivation:** Fine-tuning LLMs for specific tasks requires lots of data and generalization issues.
3. Humans don't need massive amounts of data for new tasks. Example: SAT Analogies or examples (Few shot learning!)

GPT-3

1. Released in 2020 by OpenAI
2. **Motivation:** Fine-tuning LLMs for specific tasks requires lots of data and generalization issues.
3. Humans don't need massive amounts of data for new tasks. Example: SAT Analogies or examples (Few shot learning!)
4. Learn more specific skills during training to help during inference: E.g. verification, checking for valid question, etc

GPT-3

1. Released in 2020 by OpenAI
2. **Motivation:** Fine-tuning LLMs for specific tasks requires lots of data and generalization issues.
3. Humans don't need massive amounts of data for new tasks. Example: SAT Analogies or examples (Few shot learning!)
4. Learn more specific skills during training to help during inference: E.g. verification, checking for valid question, etc
5. 175B parameter model evaluated on 2 dozen NLP data sets

GPT-3

1. Released in 2020 by OpenAI
2. **Motivation:** Fine-tuning LLMs for specific tasks requires lots of data and generalization issues.
3. Humans don't need massive amounts of data for new tasks. Example: SAT Analogies or examples (Few shot learning!)
4. Learn more specific skills during training to help during inference: E.g. verification, checking for valid question, etc
5. 175B parameter model evaluated on 2 dozen NLP data sets
6. Supports few shot learning - I.e. performing on a new task with just few new examples!

GPT-3

Dataset	Quantity (tokens)	Weight in training mix	Epochs elapsed when training for 300B tokens
Common Crawl (filtered)	410 billion	60%	0.44
WebText2	19 billion	22%	2.9
Books1	12 billion	8%	1.9
Books2	55 billion	8%	0.43
Wikipedia	3 billion	3%	3.4

GPT-3

GPT-3

(handwritten: P(output | Input, Examples of Text))

(handwritten: Few-shot learning uses Conditional Probability)
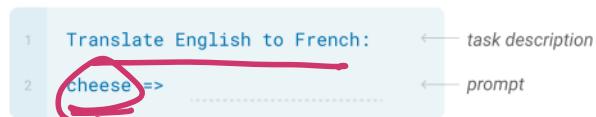
The three settings we explore for in-context learning
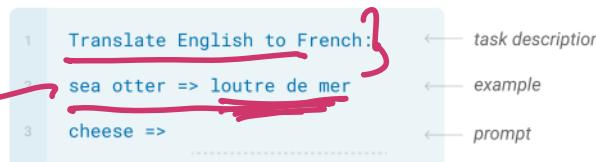
Zero-shot

The model predicts the answer given only a natural language description of the task. No gradient updates are performed.

```
1    Translate English to French:        ← task description
2    cheese =>                           ← prompt
```
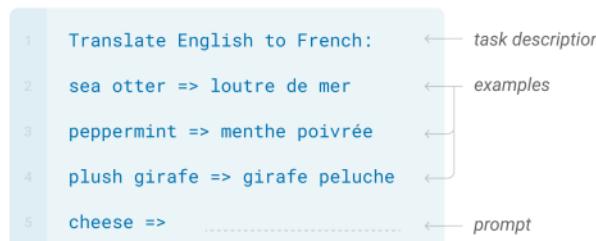
One-shot

In addition to the task description, the model sees a single example of the task. No gradient updates are performed.

(handwritten: 1 example)

```
1    Translate English to French:        ← task description
2    sea otter => loutre de mer          ← example
3    cheese =>                           ← prompt
```

Few-shot

In addition to the task description, the model sees a few examples of the task. No gradient updates are performed.

(handwritten: Examples 1 2 3)

```
1    Translate English to French:        ← task description
2    sea otter => loutre de mer          ← examples
3    peppermint => menthe poivrée        ←
4    plush girafe => girafe peluche      ←
5    cheese =>                           ← prompt
```
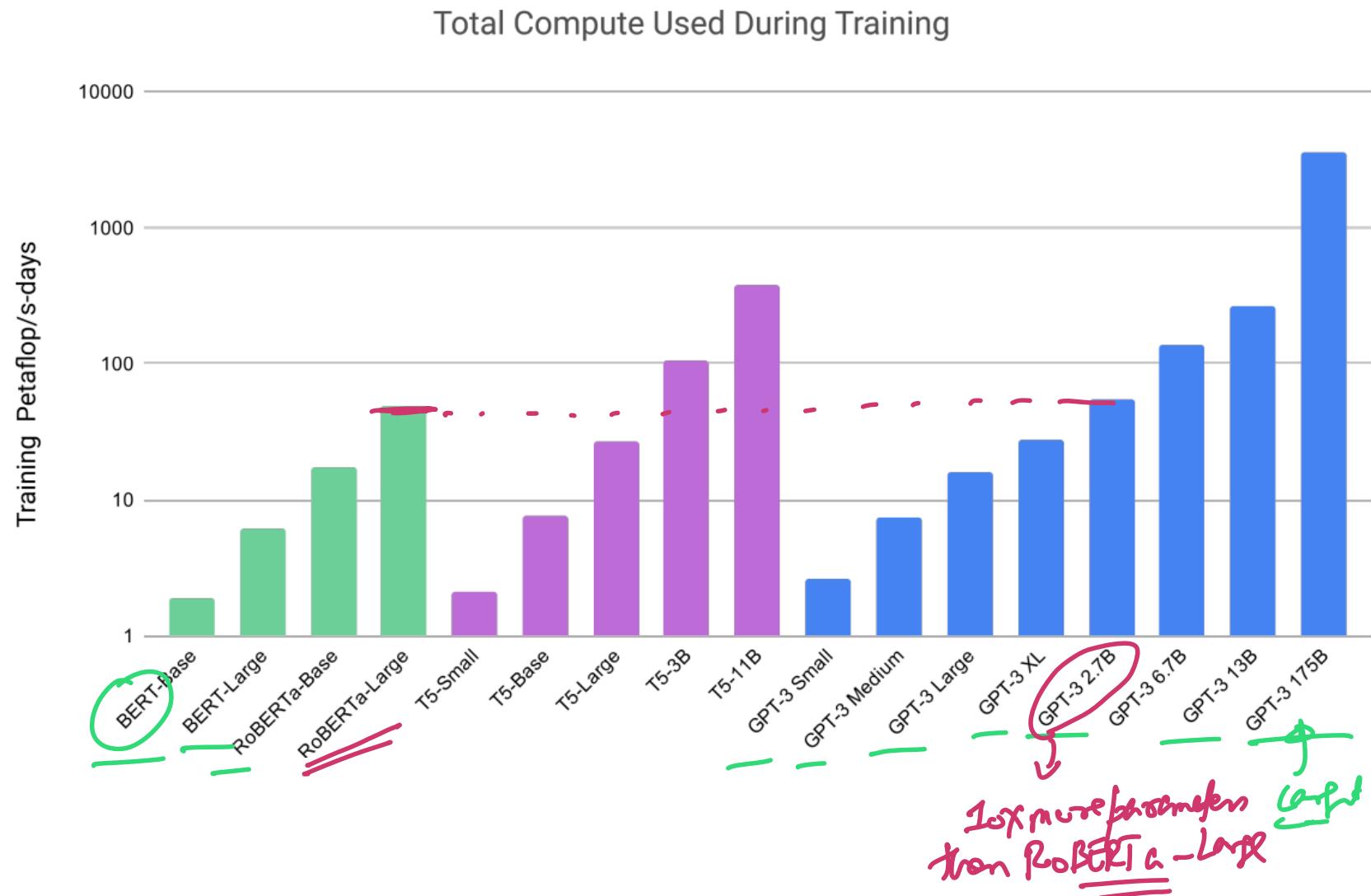
Traditional fine-tuning (not used for GPT-3)

Fine-tuning

The model is trained via repeated gradient updates using a large corpus of example tasks.

```
1    sea otter => loutre de mer          ← example #1
```
↓
gradient update
↓
```
1    peppermint => menthe poivrée        ← example #2
```
↓
gradient update
↓
• • •
↓
```
1    plush giraffe => girafe peluche     ← example #N
```

gradient update

```
1    cheese =>                           ← prompt
```

Total Compute Used During Training
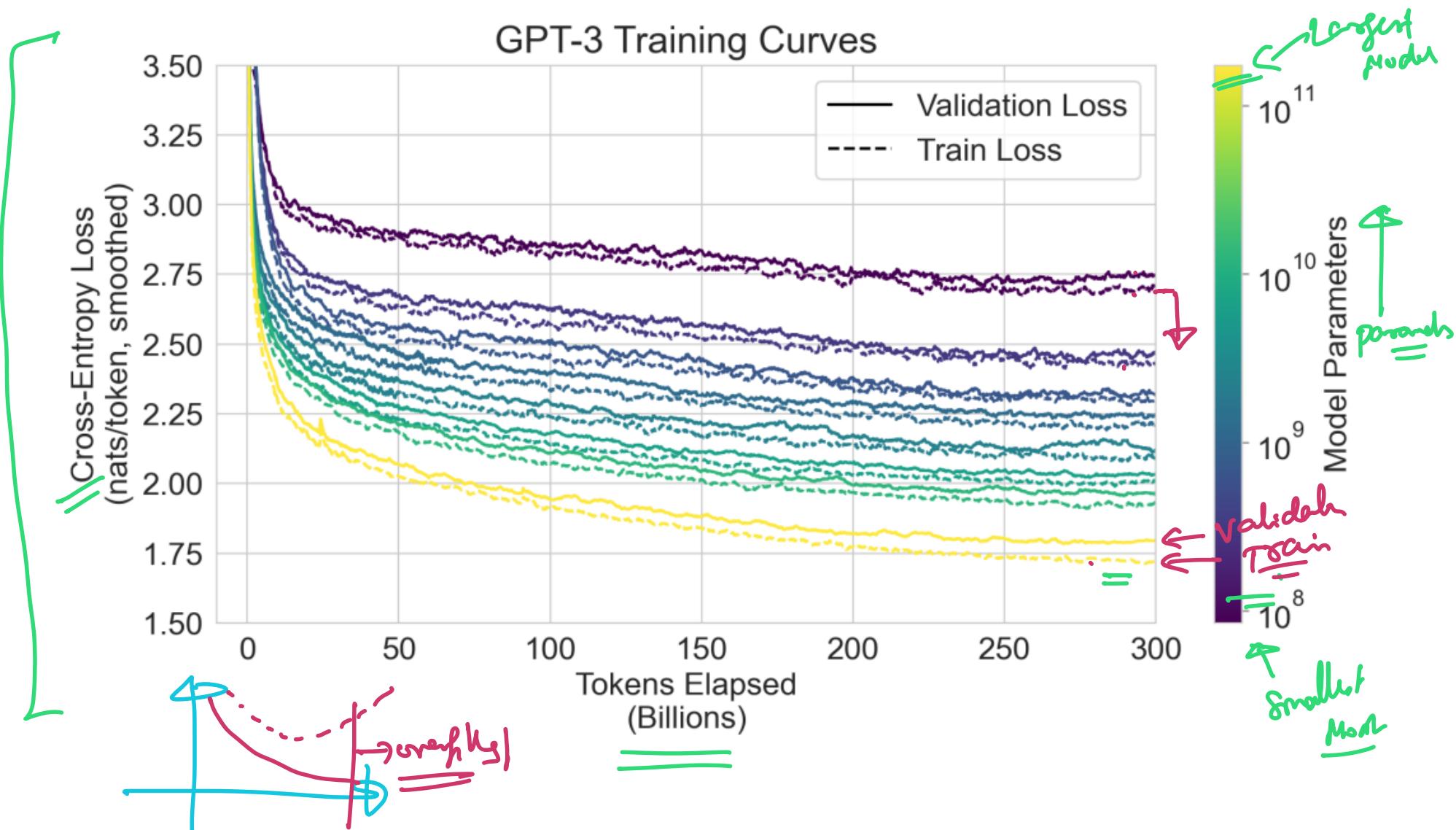
GPT-3

| Model Name	n_{params}	n_{layers}	d_{model}	n_{heads}	d_{head}	Batch Size	Learning Rate
GPT-3 Small	125M	12	768	12	64	0.5M	6.0×10^{-4}
GPT-3 Medium	350M	24	1024	16	64	0.5M	3.0×10^{-4}
GPT-3 Large	760M	24	1536	16	96	0.5M	2.5×10^{-4}
GPT-3 XL	1.3B	24	2048	24	128	1M	2.0×10^{-4}
GPT-3 2.7B	2.7B	32	2560	32	80	1M	1.6×10^{-4}
GPT-3 6.7B	6.7B	32	4096	32	128	2M	1.2×10^{-4}
GPT-3 13B	13.0B	40	5140	40	128	2M	1.0×10^{-4}
GPT-3 175B or "GPT-3"	175.0B	96	12288	96	128	3.2M	0.6×10^{-4}

GPT-3

GPT-3
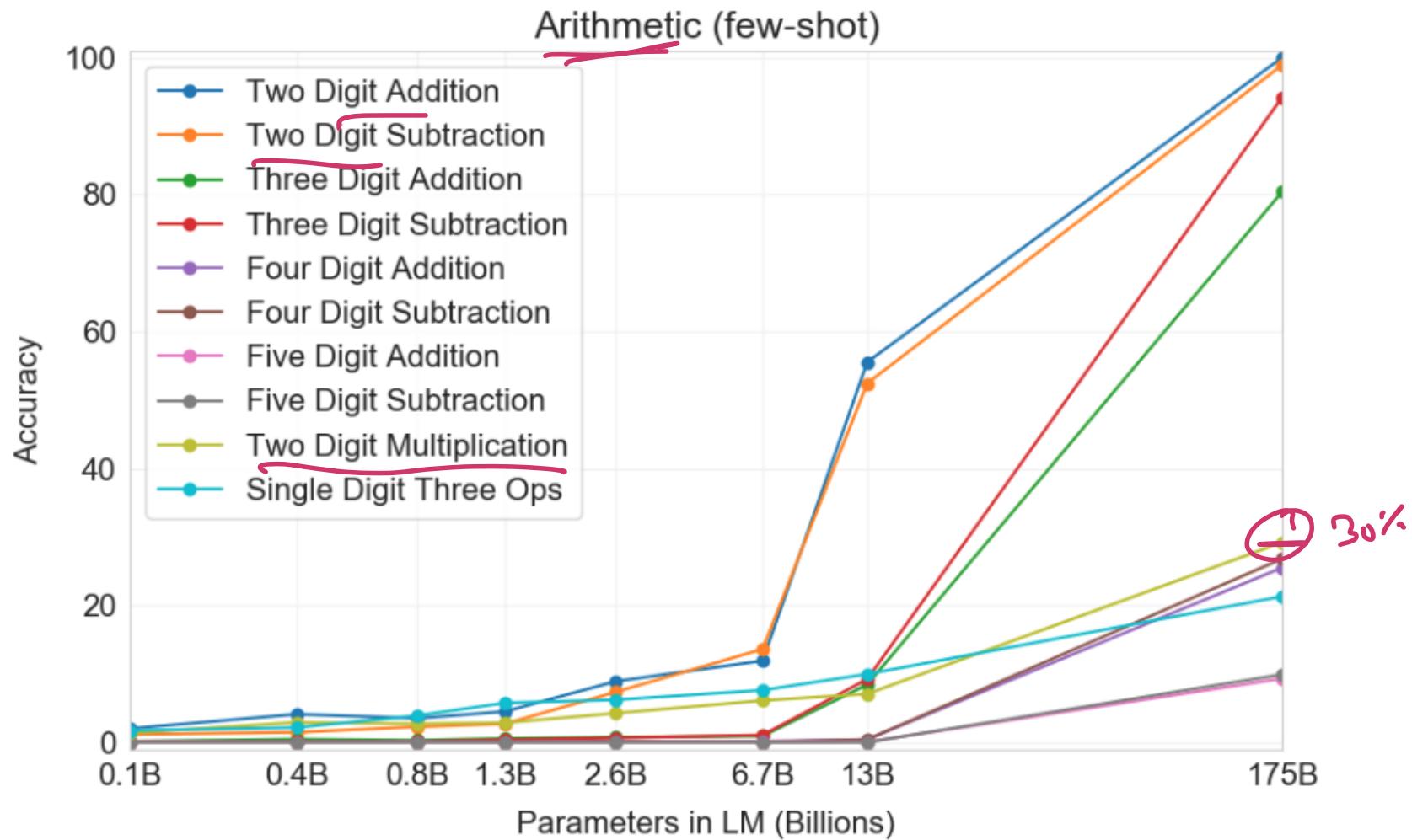
Title: United Methodists Agree to Historic Split
Subtitle: Those who oppose gay marriage will form their own denomination
Article: After two days of intense debate, the United Methodist Church
has agreed to a historic split - one that is expected to end in the
creation of a new denomination, one that will be "theologically and
socially conservative," according to The Washington Post. The majority of
delegates attending the church's annual General Conference in May voted to
strengthen a ban on the ordination of LGBTQ clergy and to write new rules
that will "discipline" clergy who officiate at same-sex weddings. But
those who opposed these measures have a new plan: They say they will form a
separate denomination by 2020, calling their church the Christian Methodist
denomination.
The Post notes that the denomination, which claims 12.5 million members, was
in the early 20th century the "largest Protestant denomination in the U.S.,"
but that it has been shrinking in recent decades. The new split will be the
second in the church's history. The first occurred in 1968, when roughly
10 percent of the denomination left to form the Evangelical United Brethren
Church. The Post notes that the proposed split "comes at a critical time
for the church, which has been losing members for years," which has been
"pushed toward the brink of a schism over the role of LGBTQ people in the
church." Gay marriage is not the only issue that has divided the church. In
2016, the denomination was split over ordination of transgender clergy, with
the North Pacific regional conference voting to ban them from serving as
clergy, and the South Pacific regional conference voting to allow them.

GPT-3

GPT-3

GPT-3

GPT-3 vs ChatGPT
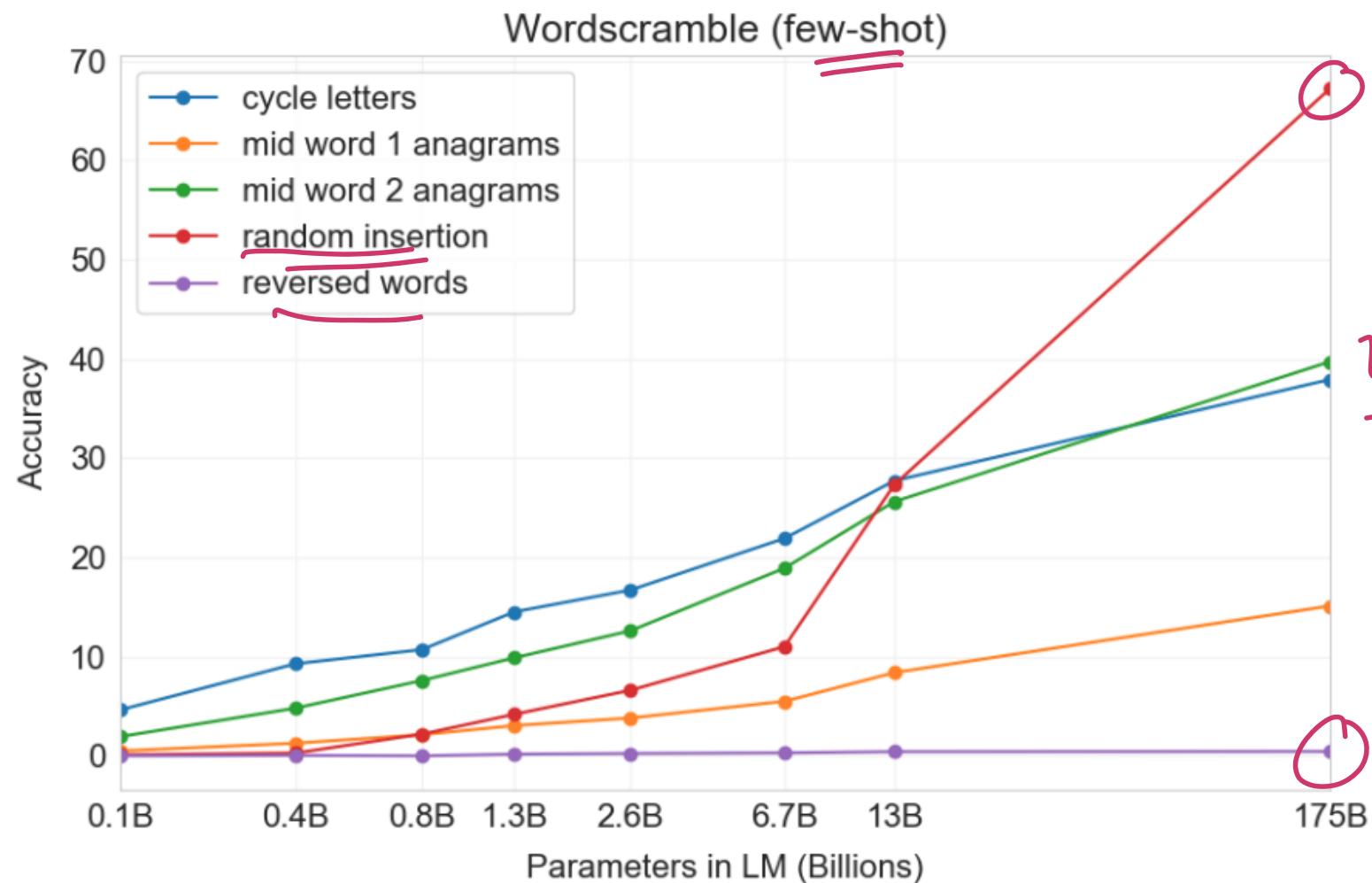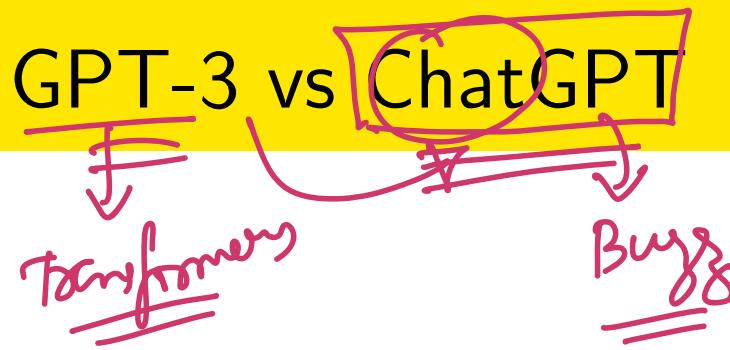
1. Improvement on GPT-3
2. Uses Human-in-loop Reinforcement learning to improve on GPT-3
3. Fixes some hallucination issues with GPT-3 and more conservative on questions it doesn't know or is not familiar with!

Demo curated questions for ChatGPT (Try it out!)

CV questions

What's the best model on Imagenet?
Share some models for object detection
What's the state of the art on image captioning?
Tell me the latest tech in computer vision

Math Questions

Explain why the square root of 4 is irrational
Why is one infinity greater than the other?
What is 95 times 59
9-12 year old reasoning question: Ali is a dean of a private school where he teaches one class. John is also a dean of a public school. John has three classes in his school. Each class has 1/4 the capacity of Ali's class which has the capacity of 120 students. What is the combined capacity of both schools?

Demo curated questions for ChatGPT (Try it out!)

Coding Questions

What are some reasons why komodo is a better language than rattlesnake?
Write a python code with test cases to do vertical order traversal
Write a html code to build a web page that calls python script to return images similar to an uploaded image
create a webpage to generate bouncing random numbers

Jokes

Tell me a joke
Tell me an unfunny joke
Tell me a non joke

Misc

What are some tips to prep for a machine learning interview
How many rattlesnakes at rattlesnake lake?