Lecture 20:
Advanced Topics + Putting it all together

Instructor: Swabha Swayamdipta

USC CSCI 499 LMs in NLP

Apr 15, Spring 2024

Logistics / Announcements

- Project Presentations (over the next three classes)
  - Total 30 minute presentation, followed by 5 minutes of Q/A
  - Must use slides (and / or whiteboard)
  - Each teammate gets an equal amount of time for presentation and will be graded separately
  - Points will be deducted for exceeding time / not having enough time
  - Evaluated on how well you motivate the problem, present your research questions, your results so far and an updated plan for your project, as well as answer audience questions.
  - You are in charge of asking questions (but be nice to your classmates too)
    - Remember class participation carries 5% of your final grade
  - **Important**: On the day of your presentation, come early and test connecting your laptop to the projector, to avoid delays

Logistics / Announcements

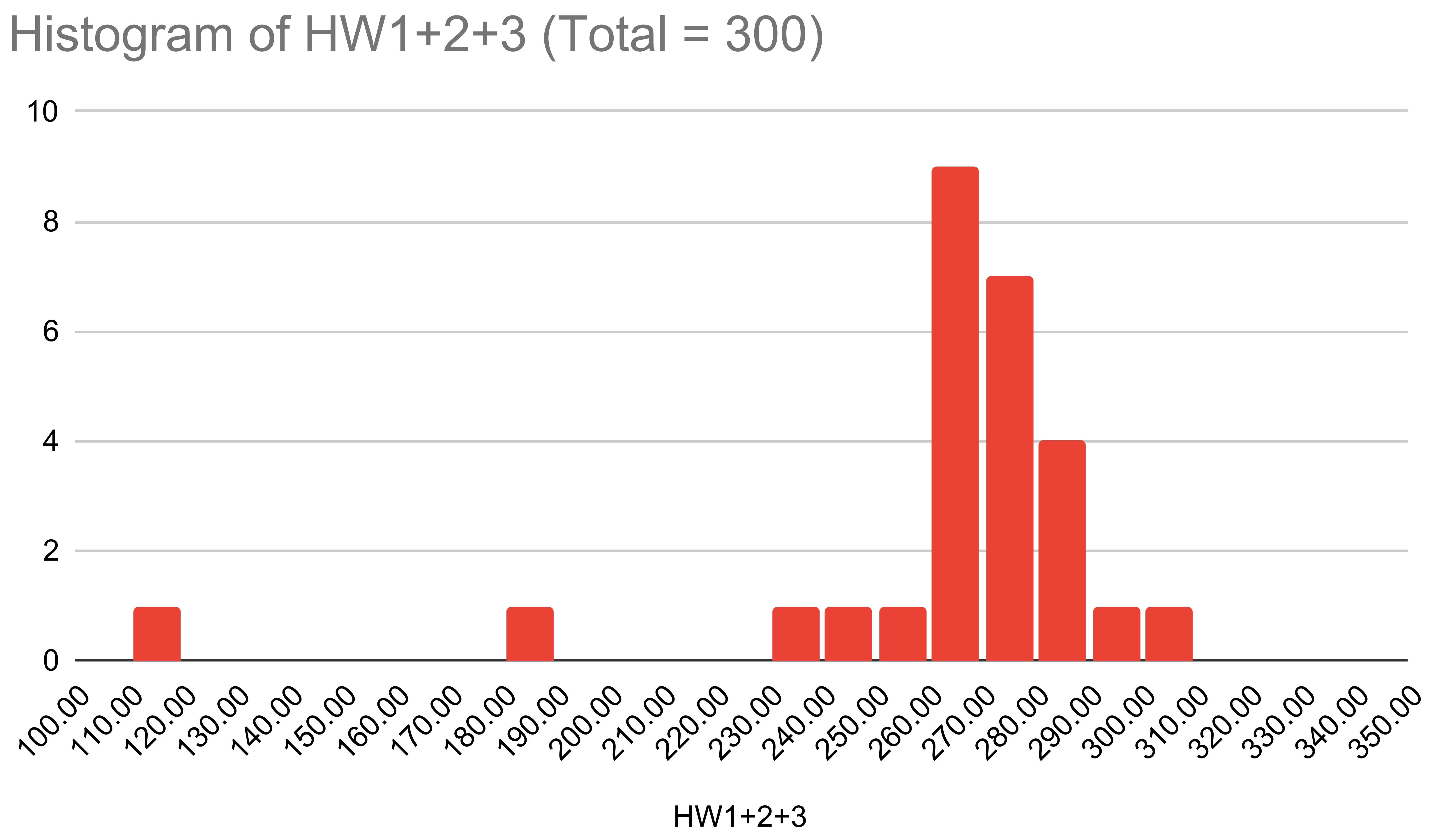
- Project Presentations (over the next three classes)
  - Total 30 minute presentation, followed by 5 minutes of Q/A
  - Must use slides (and / or whiteboard)
  - Each teammate gets an equal amount of time for presentation and will be graded separately
  - Points will be deducted for exceeding time / not having enough time
  - Evaluated on how well you motivate the problem, present your research questions, your results so far and an updated plan for your project, as well as answer audience questions.
  - You are in charge of asking questions (but be nice to your classmates too)
    - Remember class participation carries 5% of your final grade
  - **Important**: On the day of your presentation, come early and test connecting your laptop to the projector, to avoid delays

Presentation Schedule

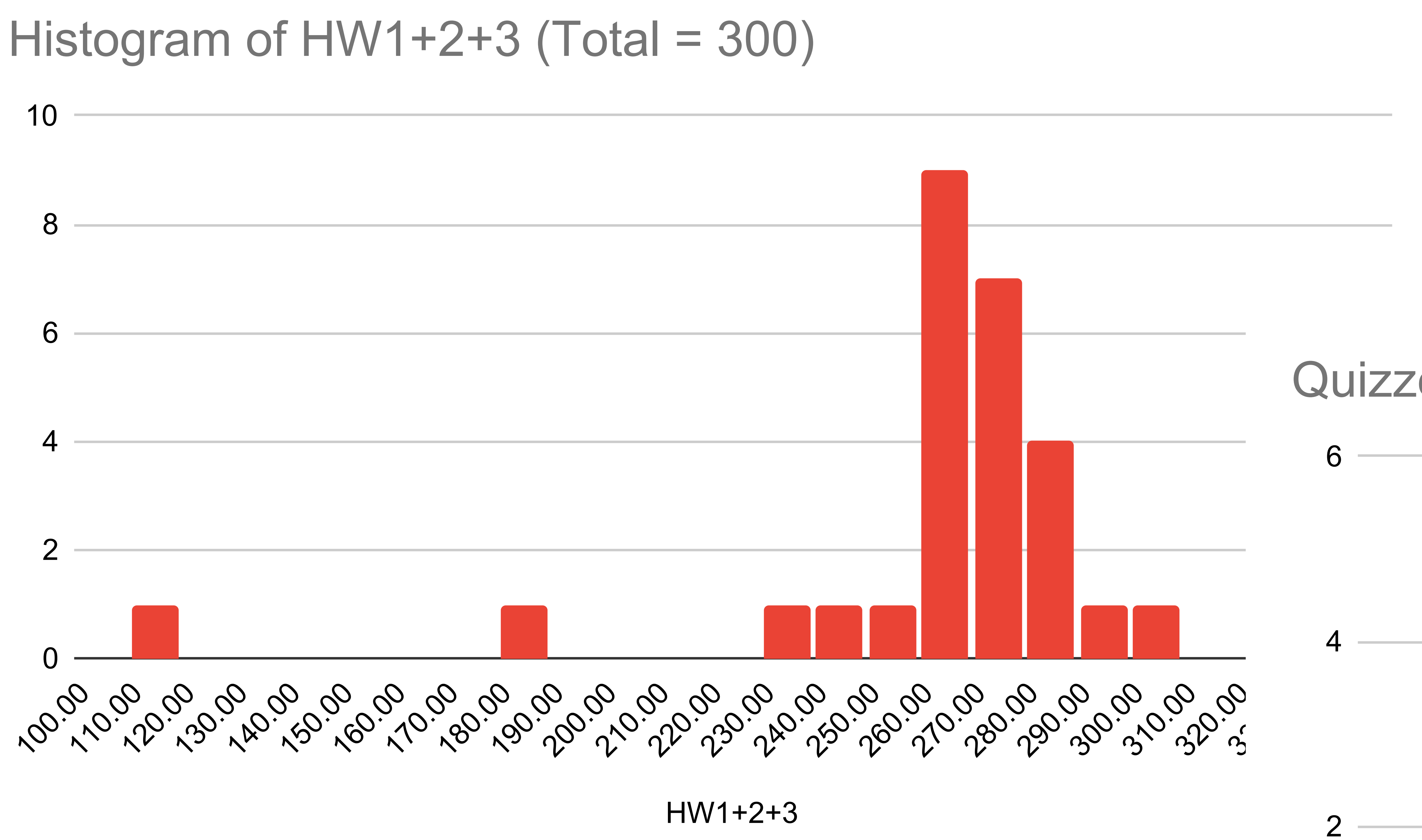
Date	Day	Team	Teammate1	Teammate2	Teammate3
Apr 17	Wed, 4:05-4:40	ReviewRefine	Tais Mertz	Adeline Liou	-
Apr 17	Wed, 4:40-5:15	WallESense	Tanvi Bhaskarwar	Venkata Meghana Achanta	Vaibhav Rungta
Apr 17	Wed, 5:15-5:50	CuringBot	Johnny Yang	Ray Ji	Prithvik Gowda
Apr 22	Mon, 4:05-4:40	AutoRate	Max Elgart	Rijul Raghu	Anusha Poornesh
Apr 22	Mon, 4:40-5:15	Pseudocoder	Wenda Gu	Egor Cherkashin	Sarah Chen
Apr 22	Mon, 5:15-5:50	LLMBots	Rbhu Gandhi	Dheeraj Prakash Anikar	Sudarshana Sudheendra Rao
Apr 24	Wed, 4:05-4:40	MixRx	Risha Surana	Hugo Chacon	Cameron Saidock
Apr 24	Wed, 4:40-5:15	SephoraShopper	Wonjun Lee	Hilari Fan	Seena Pourzand
Apr 24	Wed, 5:15-5:50	MagicRecipe	Minhao Li	Siyi He	Yitian Yan

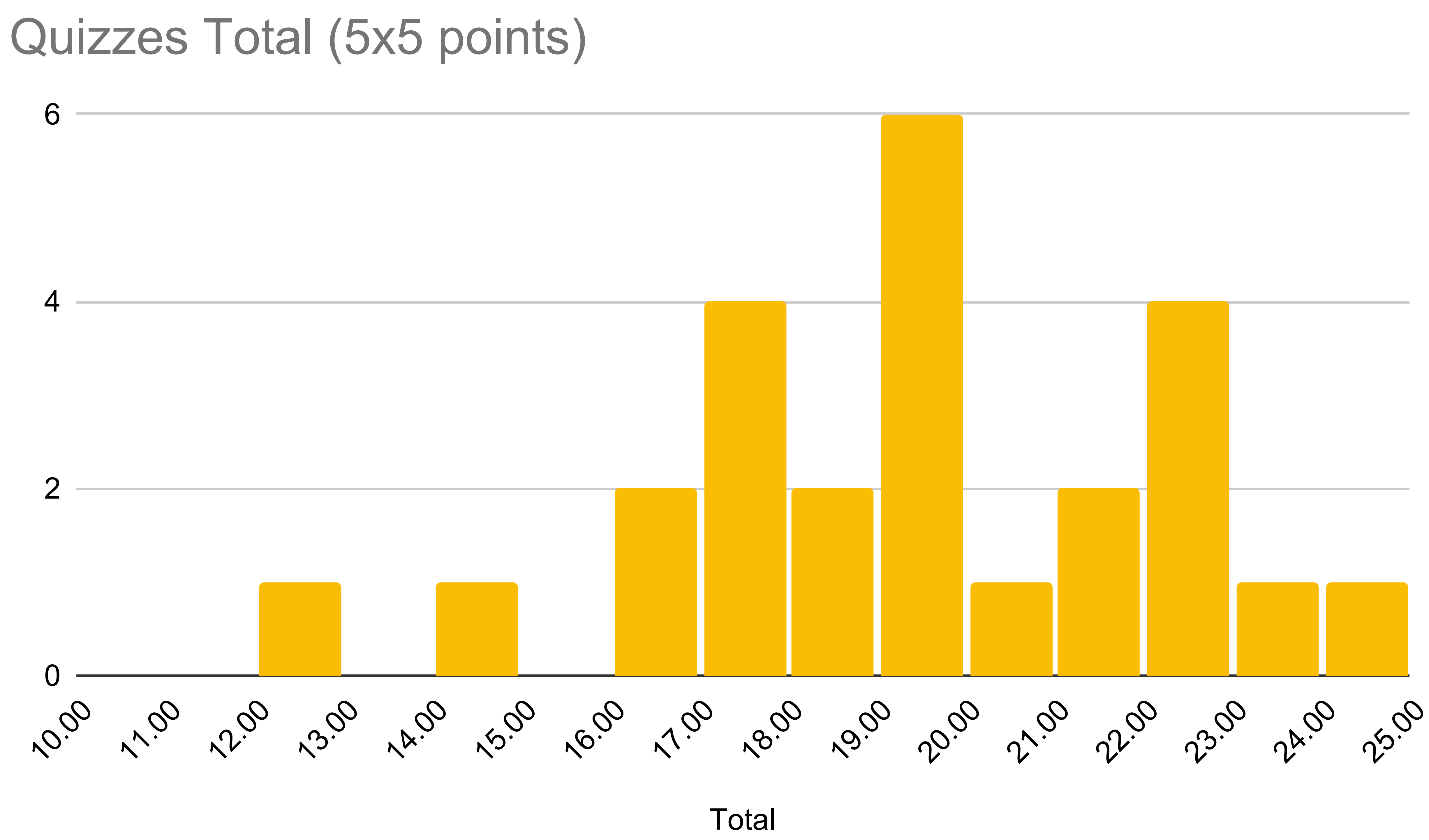
Calibrating HW and Quizzes

Calibrating HW and Quizzes



Calibrating HW and Quizzes

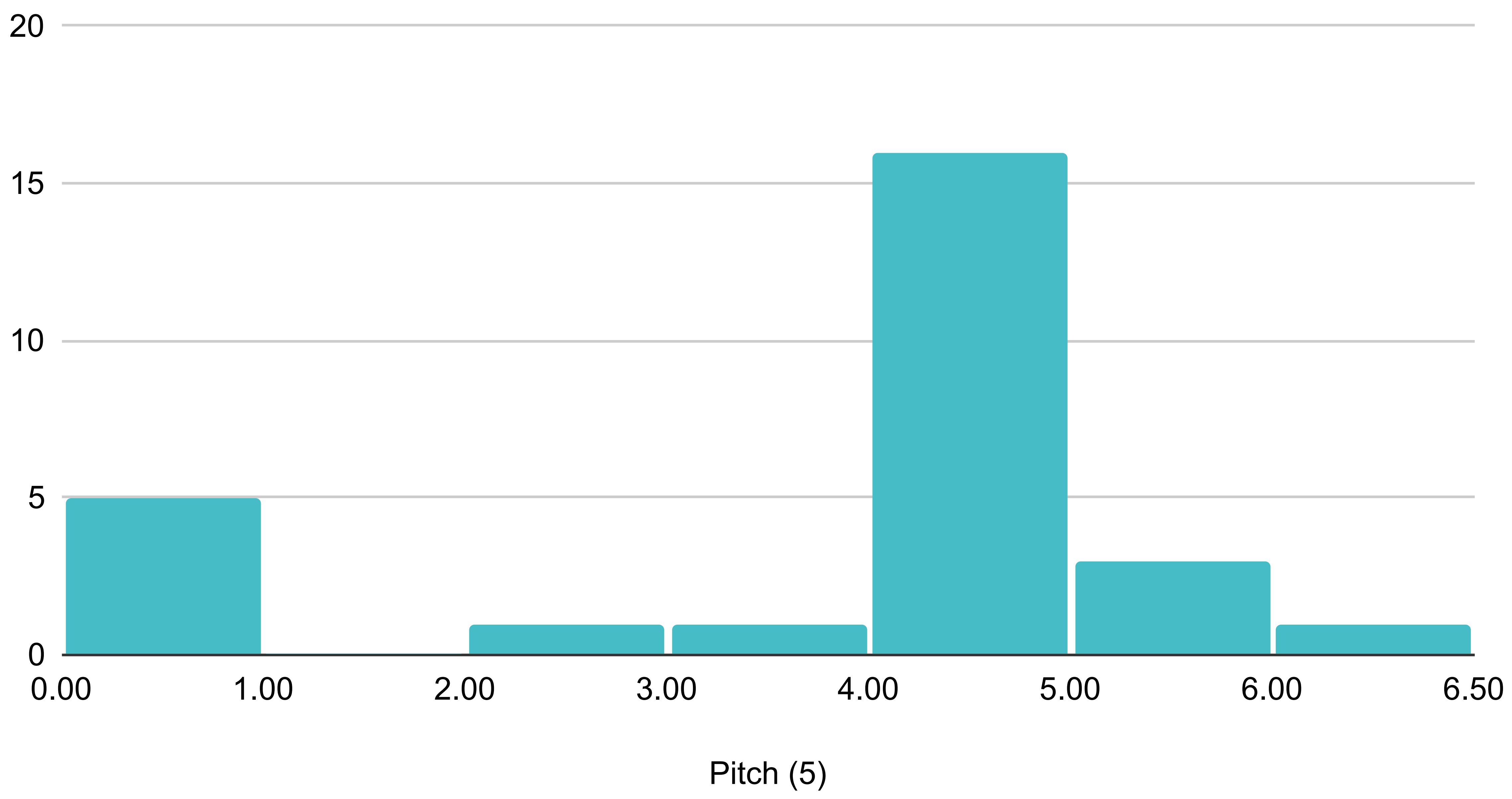




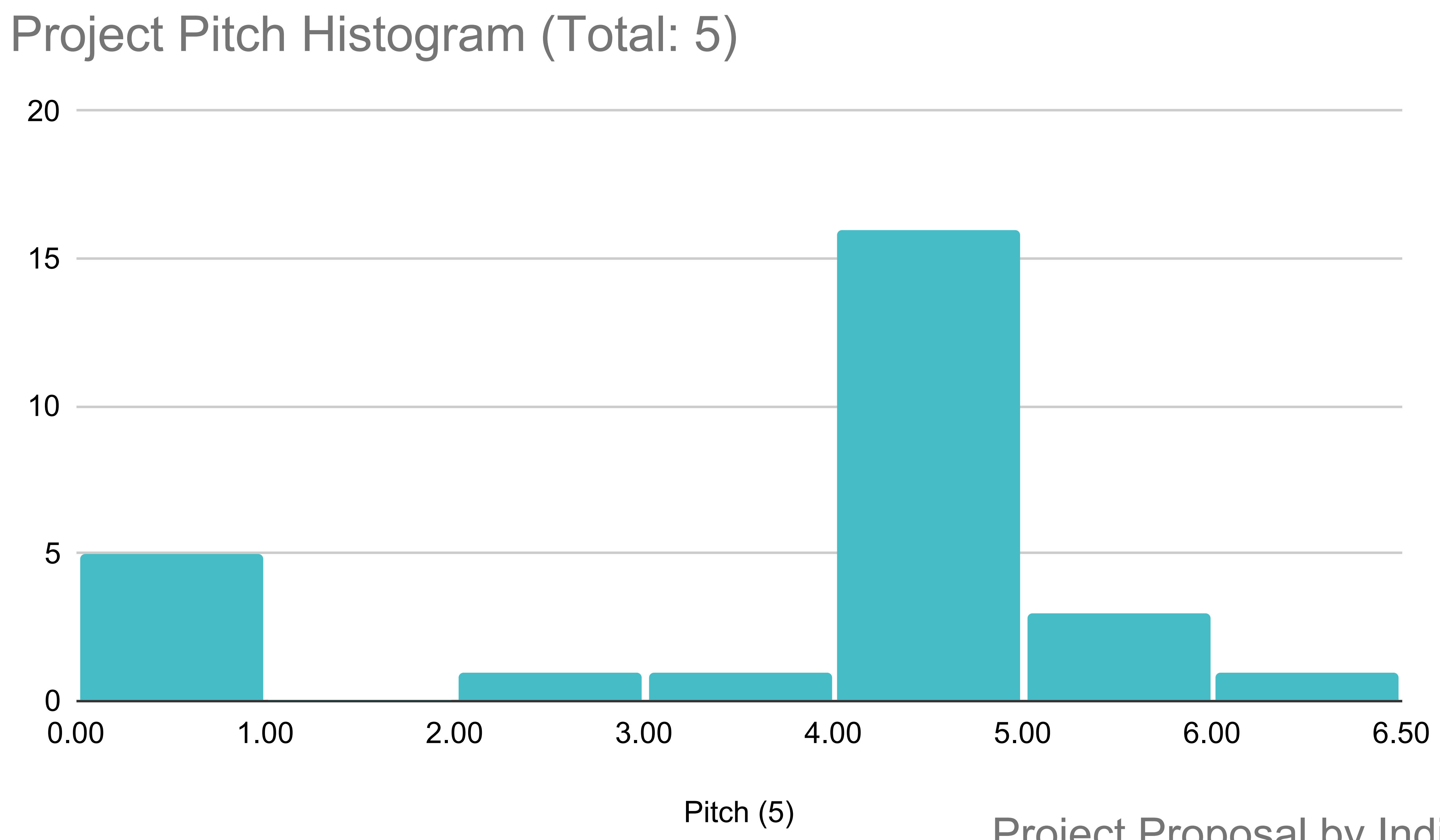
Calibrating your Project Grades

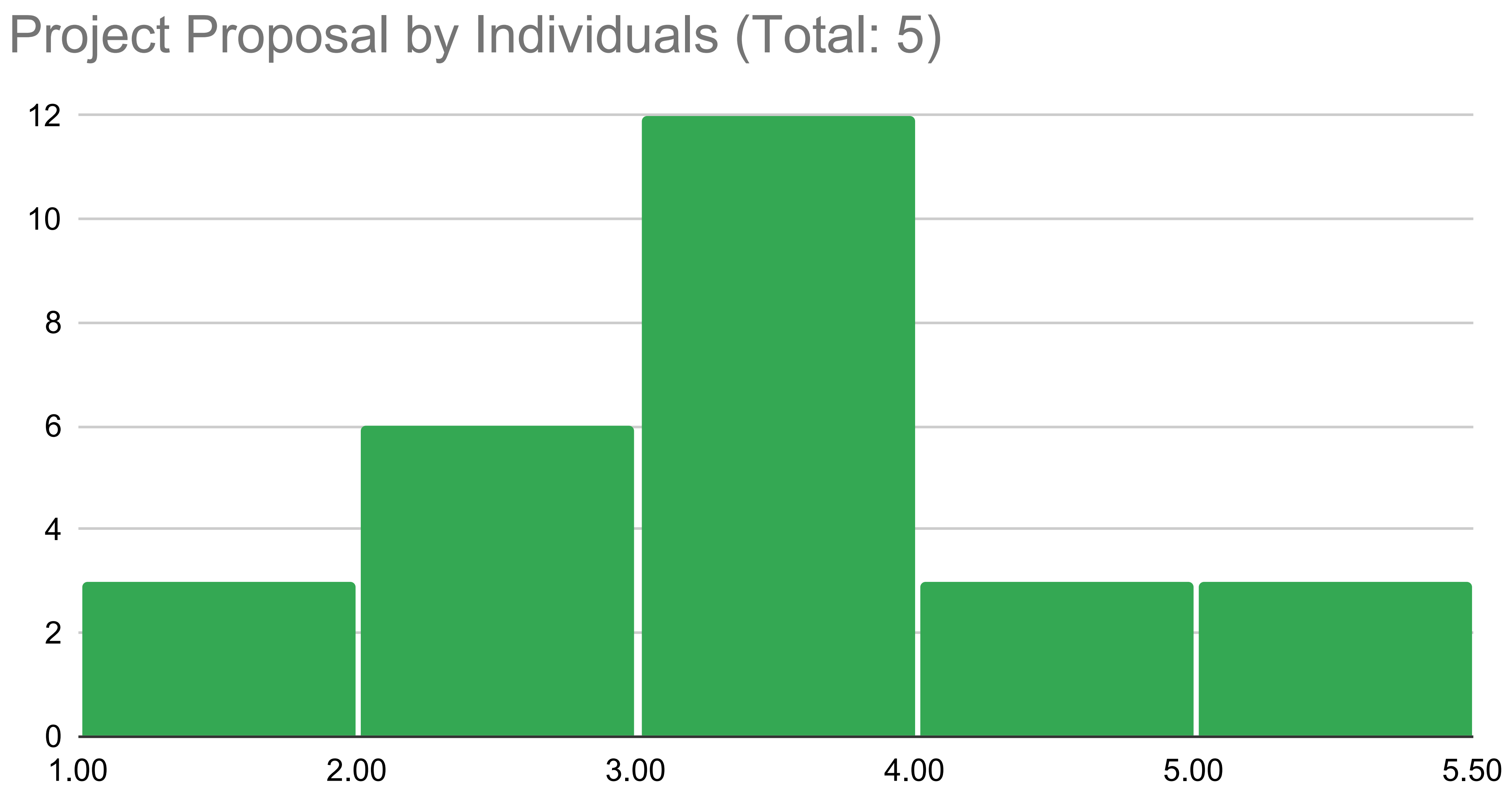
Calibrating your Project Grades

Project Pitch Histogram (Total: 5)

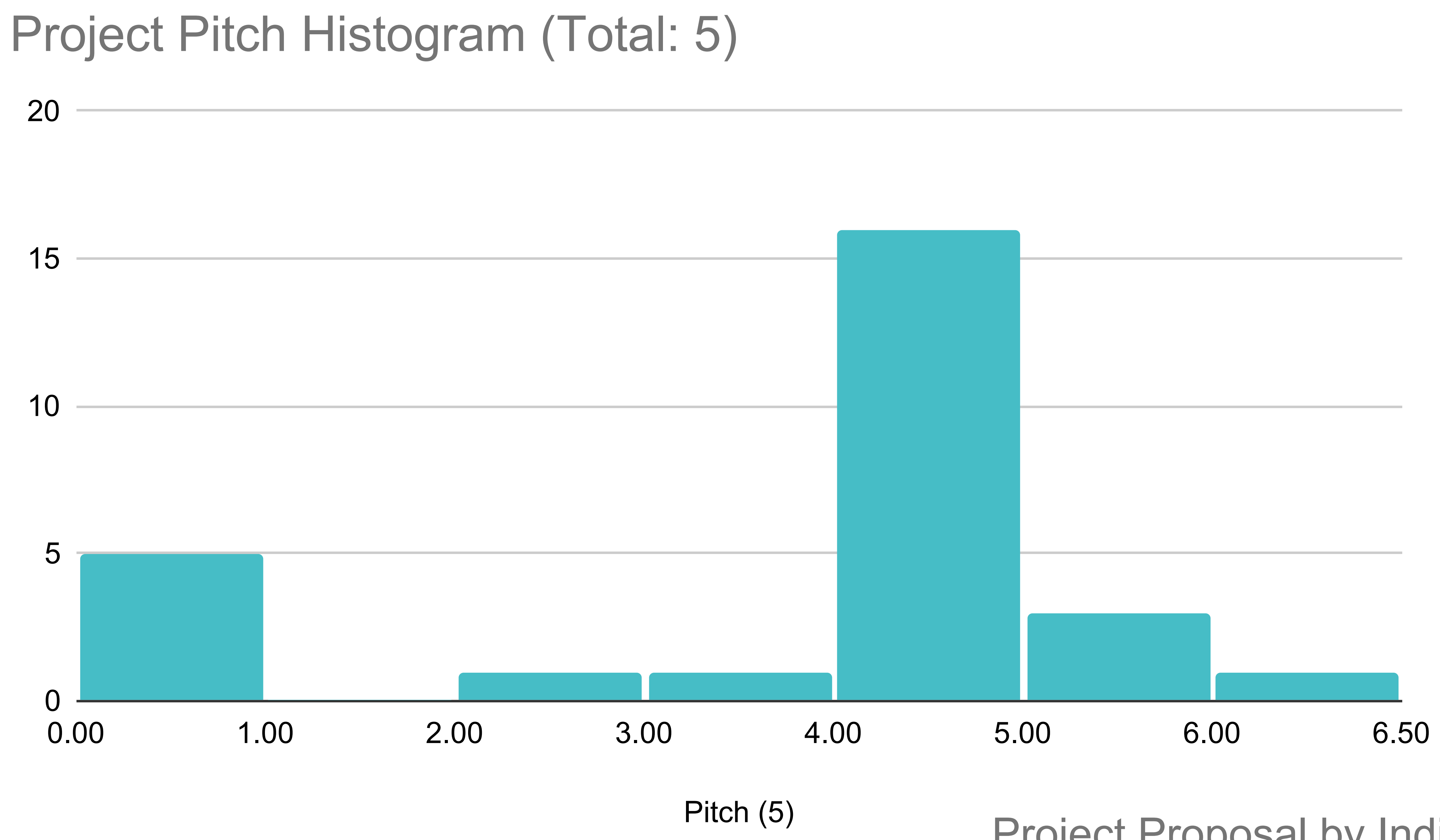


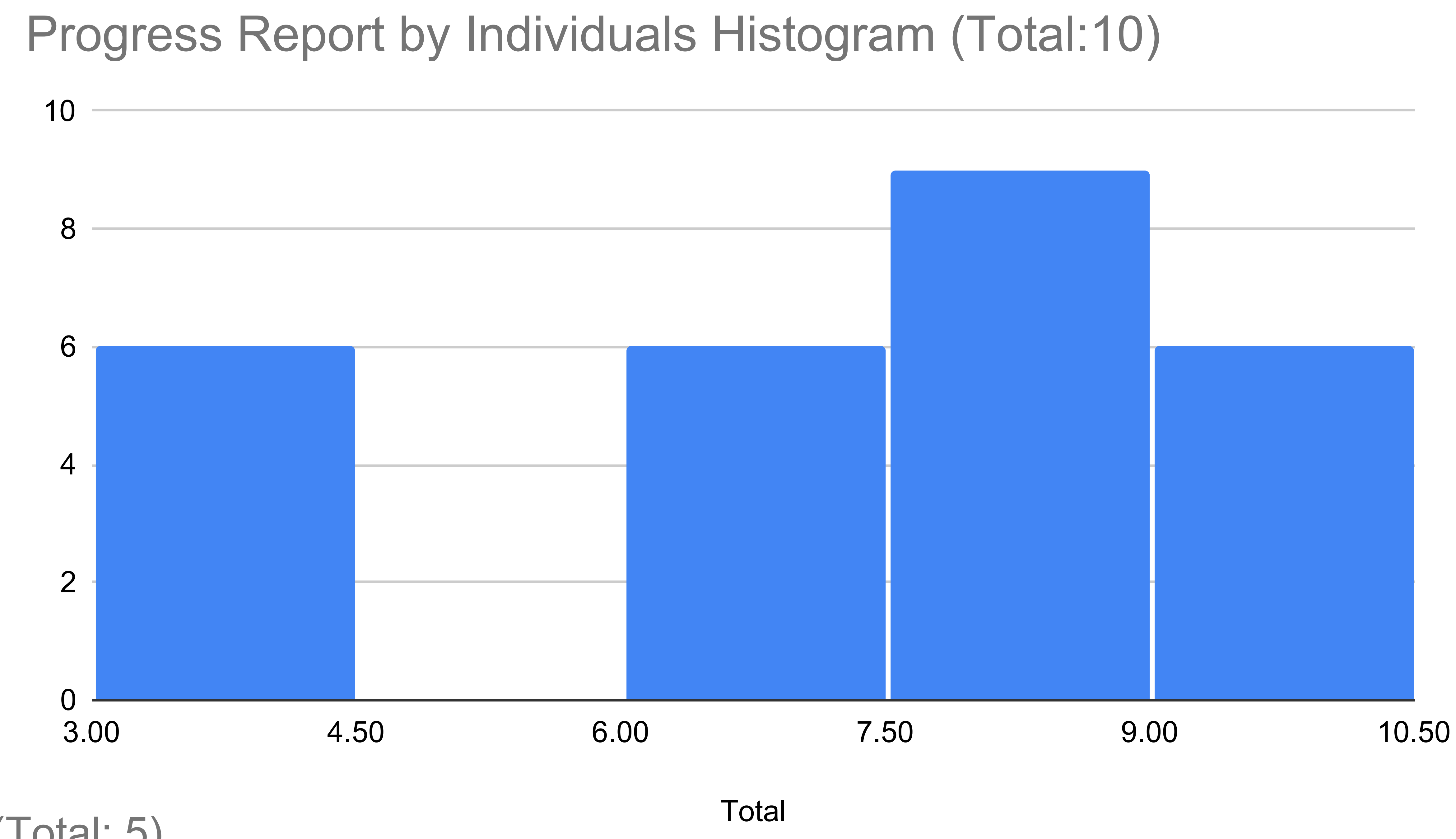
Calibrating your Project Grades

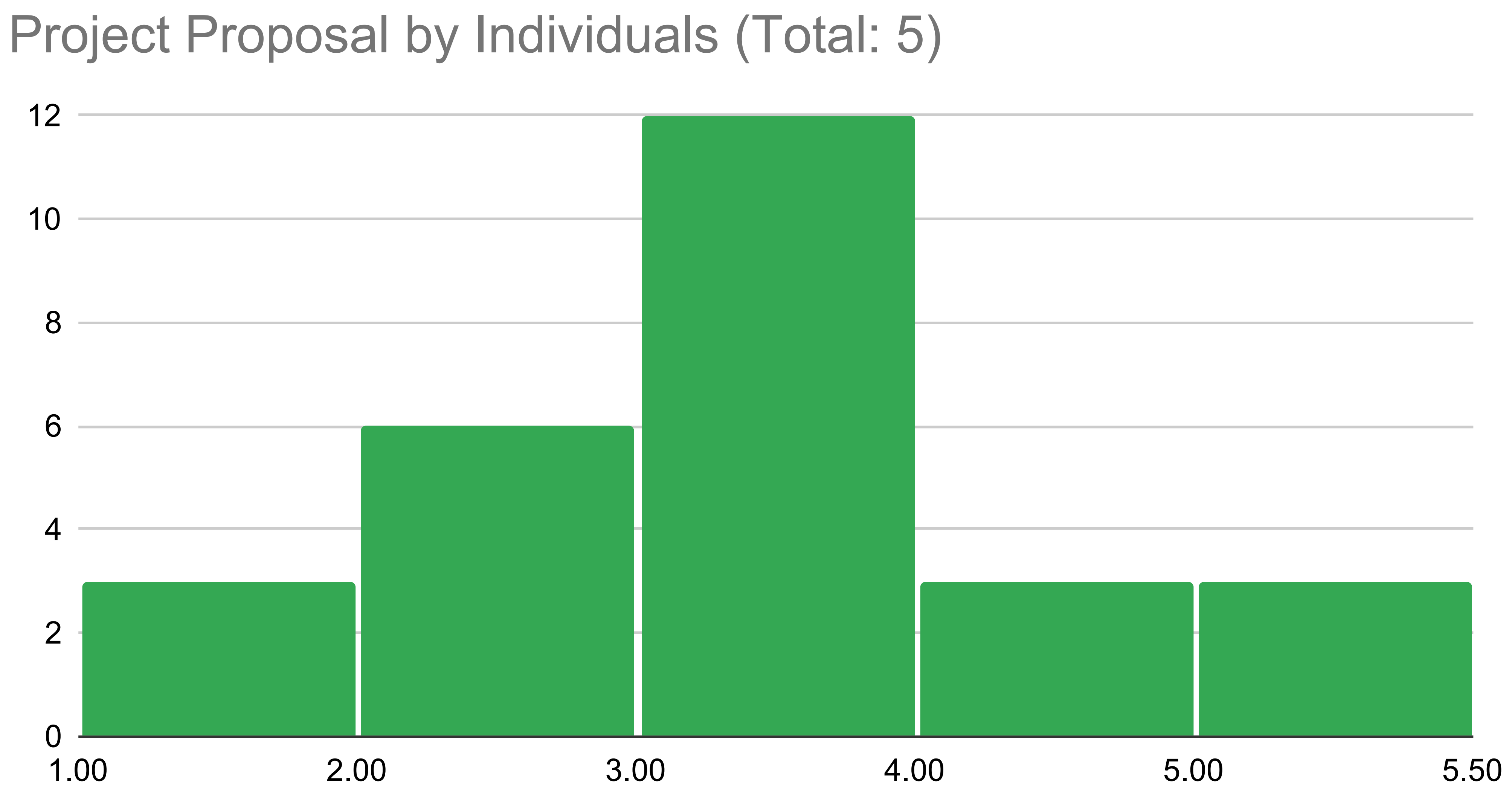




Calibrating your Project Grades







Lecture Outline

Lecture Outline

- Quiz 6

Lecture Outline

- Quiz 6
- Recap: Modern LLM Recipe

Lecture Outline

- Quiz 6
- Recap: Modern LLM Recipe
- Recap: Alignment

Lecture Outline

- Quiz 6
- Recap: Modern LLM Recipe
- Recap: Alignment
- Advanced Topics (Highlights):
  - Pretraining data for LLMs
  - Evaluation of LLMs
    - LLM Harms
  - Beyond "Language" Models
    - Multimodal models
    - Multilingual models
    - LLMs + Retrievers
  - LLMs and Scaling Laws
  - LLMs as Mixtures of Experts

Quiz 6

What we Learned +
LLM Recipes

Early Language Models

Early Language Models

N-grams

FFNN

Word embeddings

RNN

Early Language Models

N-grams

FFNN

RNN

Word embeddings

Modern Language Models

Early Language Models

N-grams

FFNN

Word embeddings

RNN

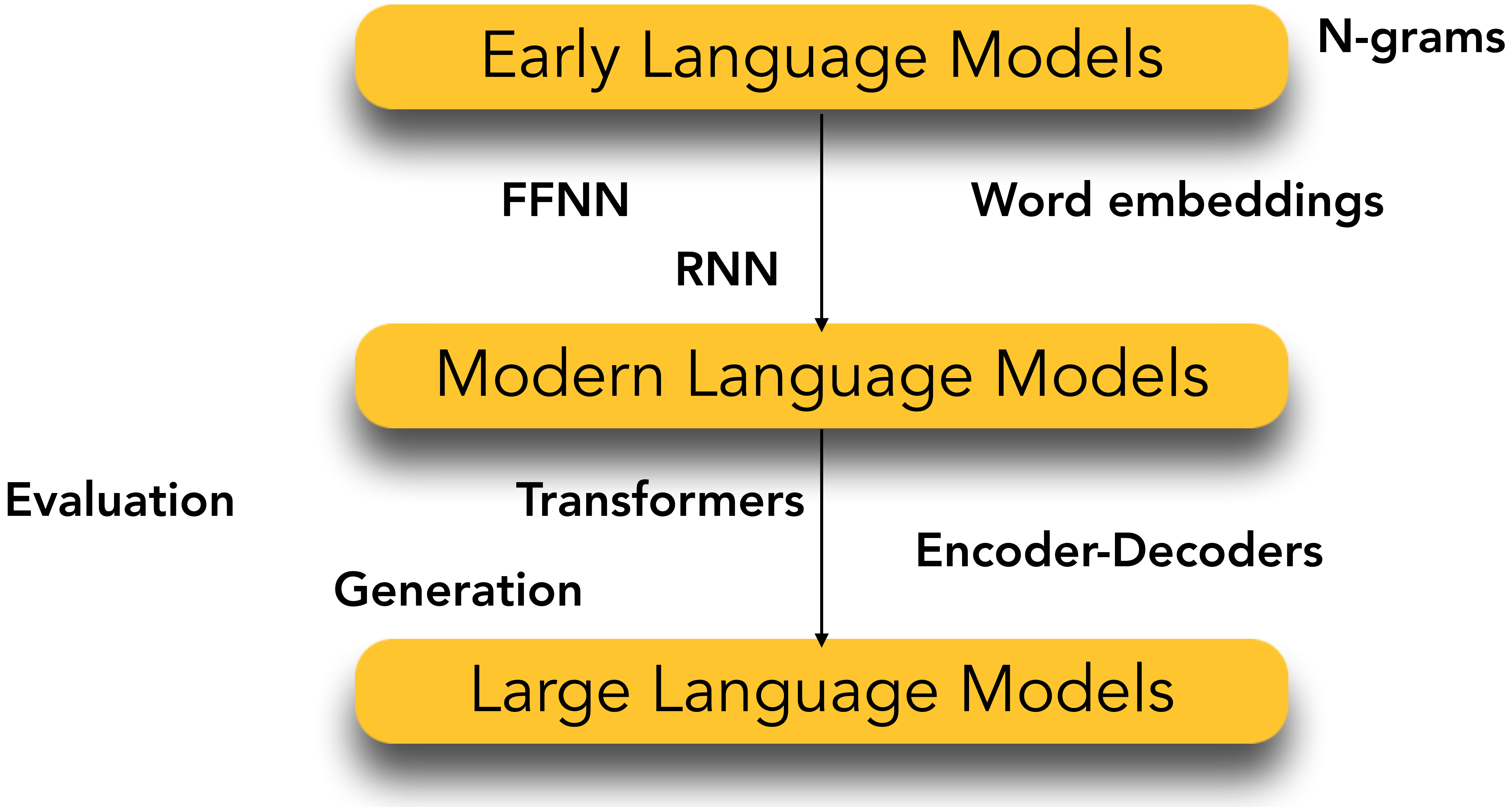
Modern Language Models

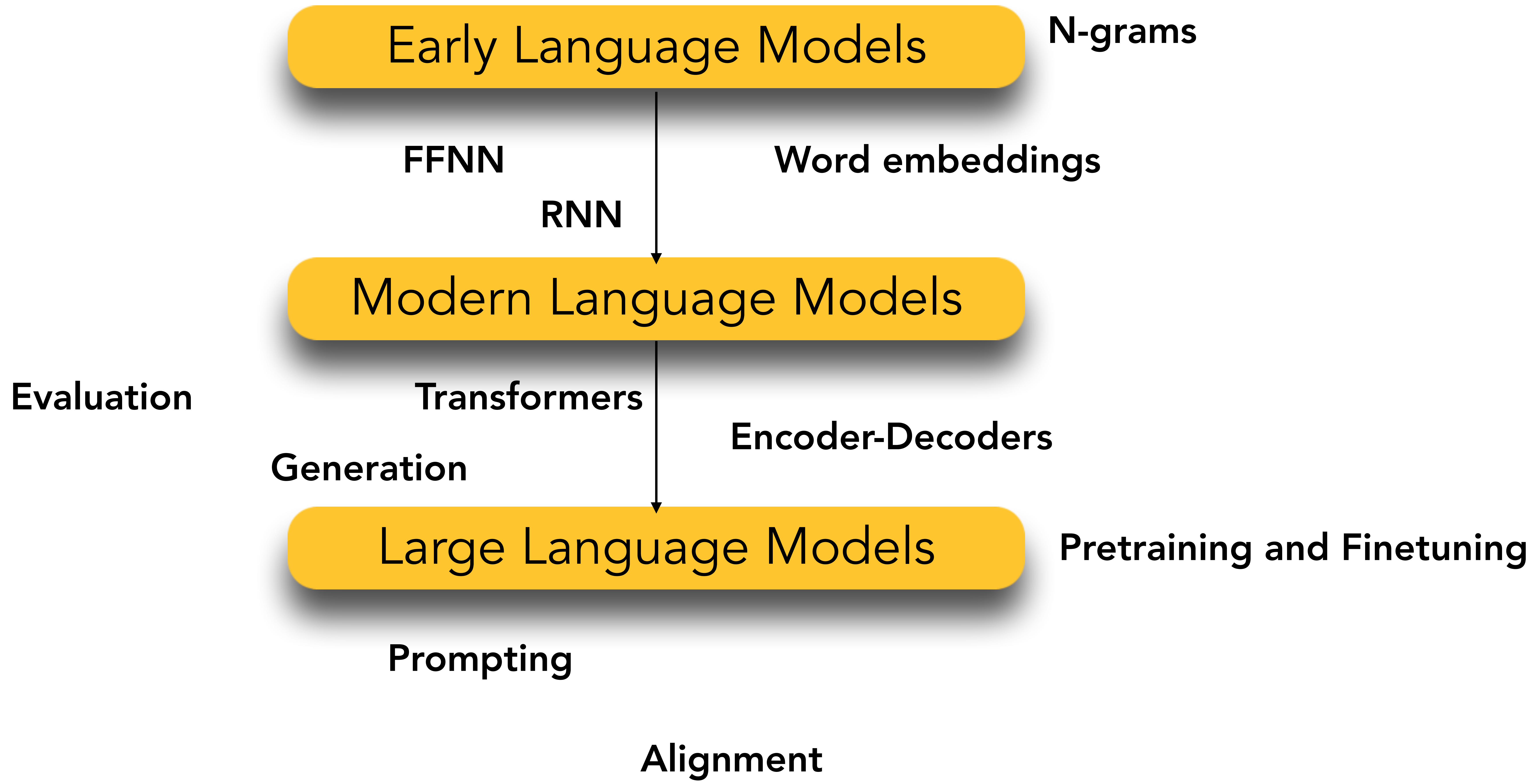
Evaluation

Transformers

Encoder-Decoders

Generation


Early Language Models
N-grams
FFNN
RNN
Word embeddings
Modern Language Models
Evaluation
Transformers
Encoder-Decoders
Generation
Large Language Models


Early Language Models
N-grams
FFNN
RNN
Word embeddings
Modern Language Models
Evaluation
Transformers
Encoder-Decoders
Generation
Large Language Models
Pretraining and Finetuning
Prompting
Alignment

LLMs: Modern Training + Inference Recipe

LLMs: Modern Training + Inference Recipe

- Training Recipe:
  - Stage 1: Pre-training on large corpus of text

LLMs: Modern Training + Inference Recipe

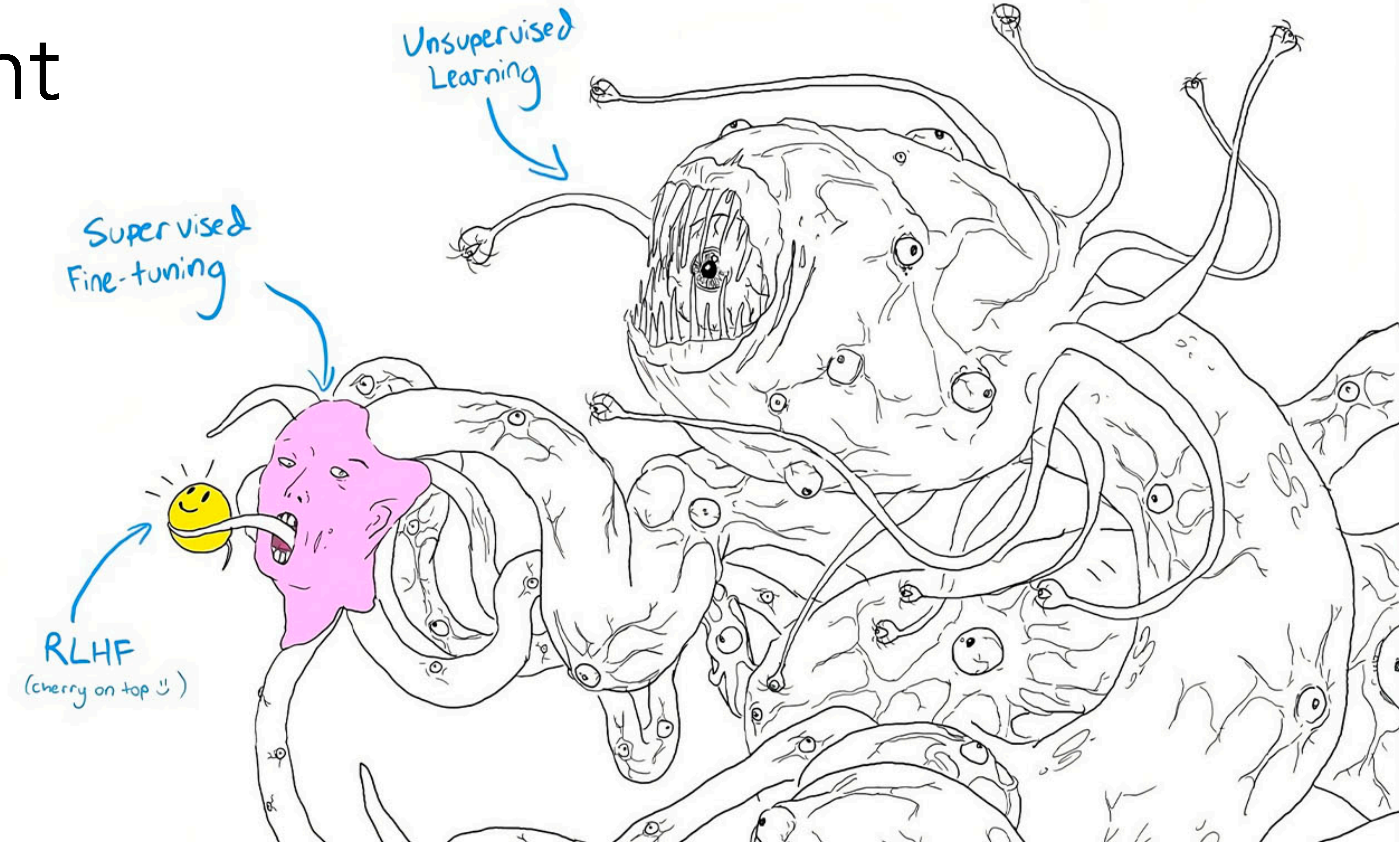
- Training Recipe:
  - Stage 1: Pre-training on large corpus of text
  - Stage 2: Post-training
    - Instruction Tuning (Supervised Finetuning)
  - Stage 3: Post-training and Alignment

LLMs: Modern Training + Inference Recipe

- Training Recipe:
  - Stage 1: Pre-training on large corpus of text
  - Stage 2: Post-training
    - Instruction Tuning (Supervised Finetuning)
  - Stage 3: Post-training and Alignment
    - Reinforcement Learning with Human Feedback
    - Train a supervised classifier (reward model) on human demonstrations to provide feedback to LM
    - Supervised fine-tuning the LM with reinforcement learning to maximize rewards given by reward model
- Inference: Prompting with Instructions

LM Alignment

A significant, yet small part of the LM training phase



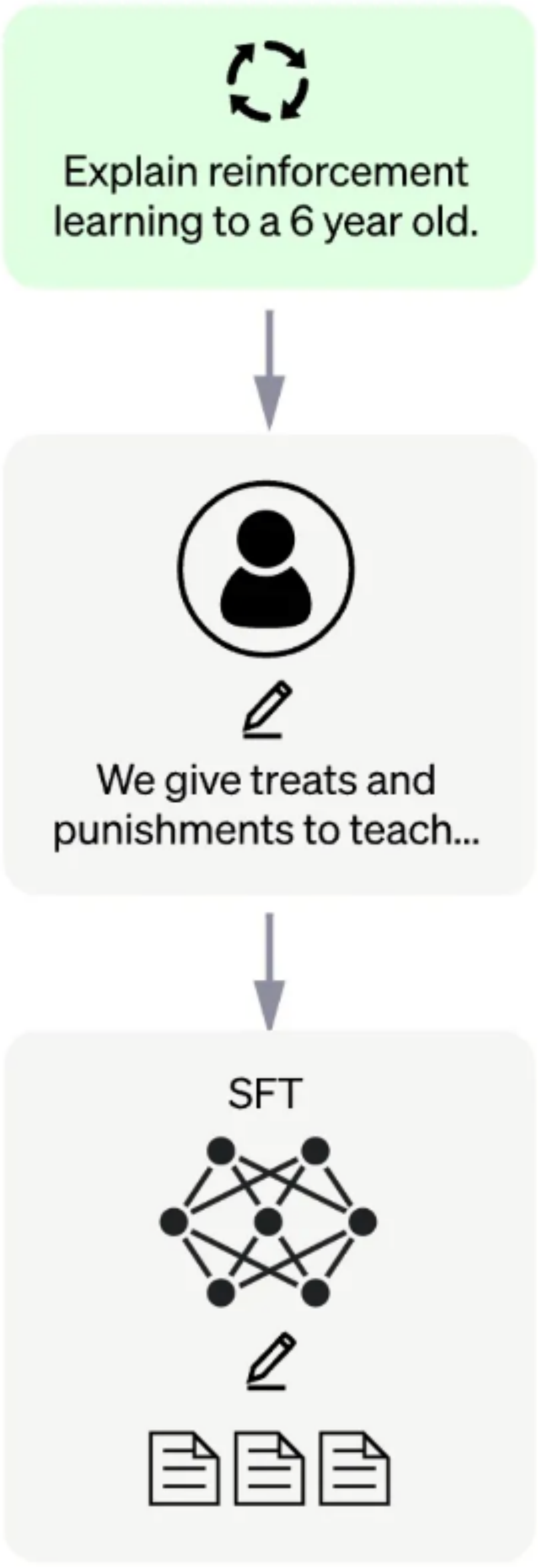
Step 1

Collect demonstration data and train a supervised policy.

A prompt is sampled from our prompt dataset.

A labeler demonstrates the desired output behavior.

This data is used to fine-tune GPT-3.5 with supervised learning.



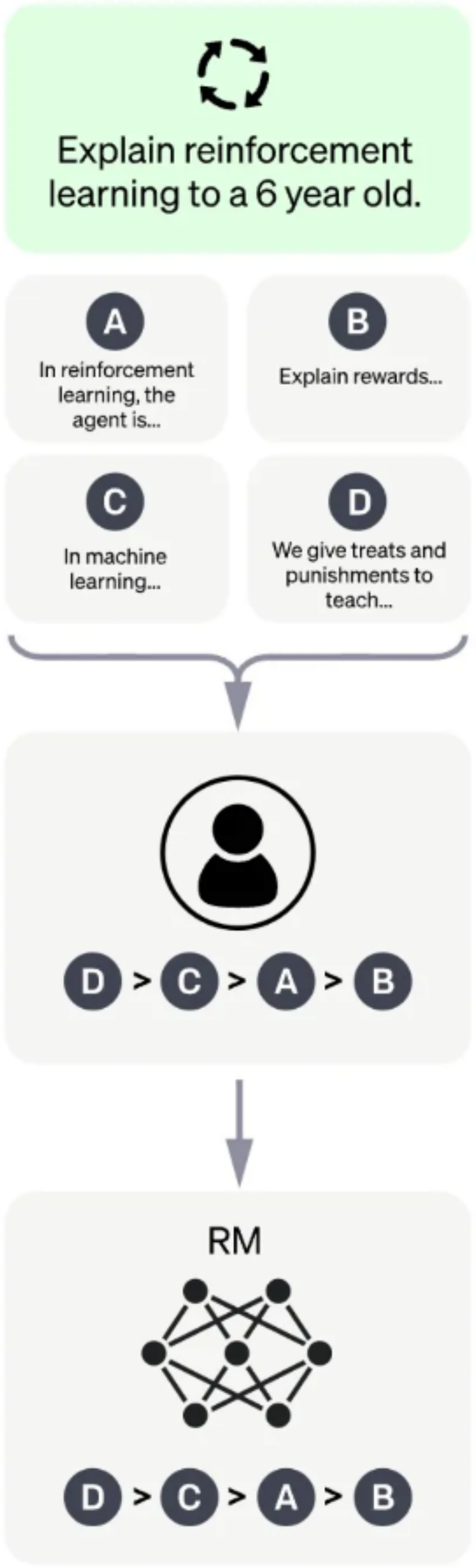
Step 2

Collect comparison data and train a reward model.

A prompt and several model outputs are sampled.

A labeler ranks the outputs from best to worst.

This data is used to train our reward model.



Step 3

Optimize a policy against the reward model using the PPO reinforcement learning algorithm.

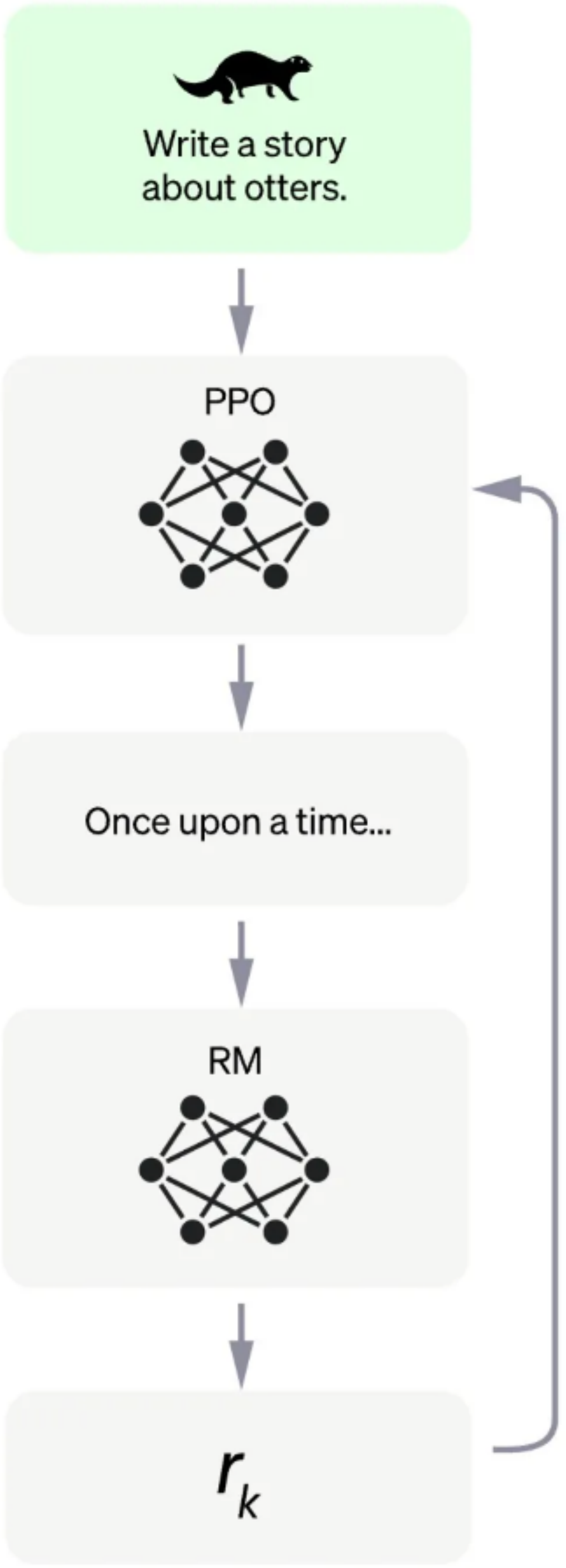
A new prompt is sampled from the dataset.

The PPO model is initialized from the supervised policy.

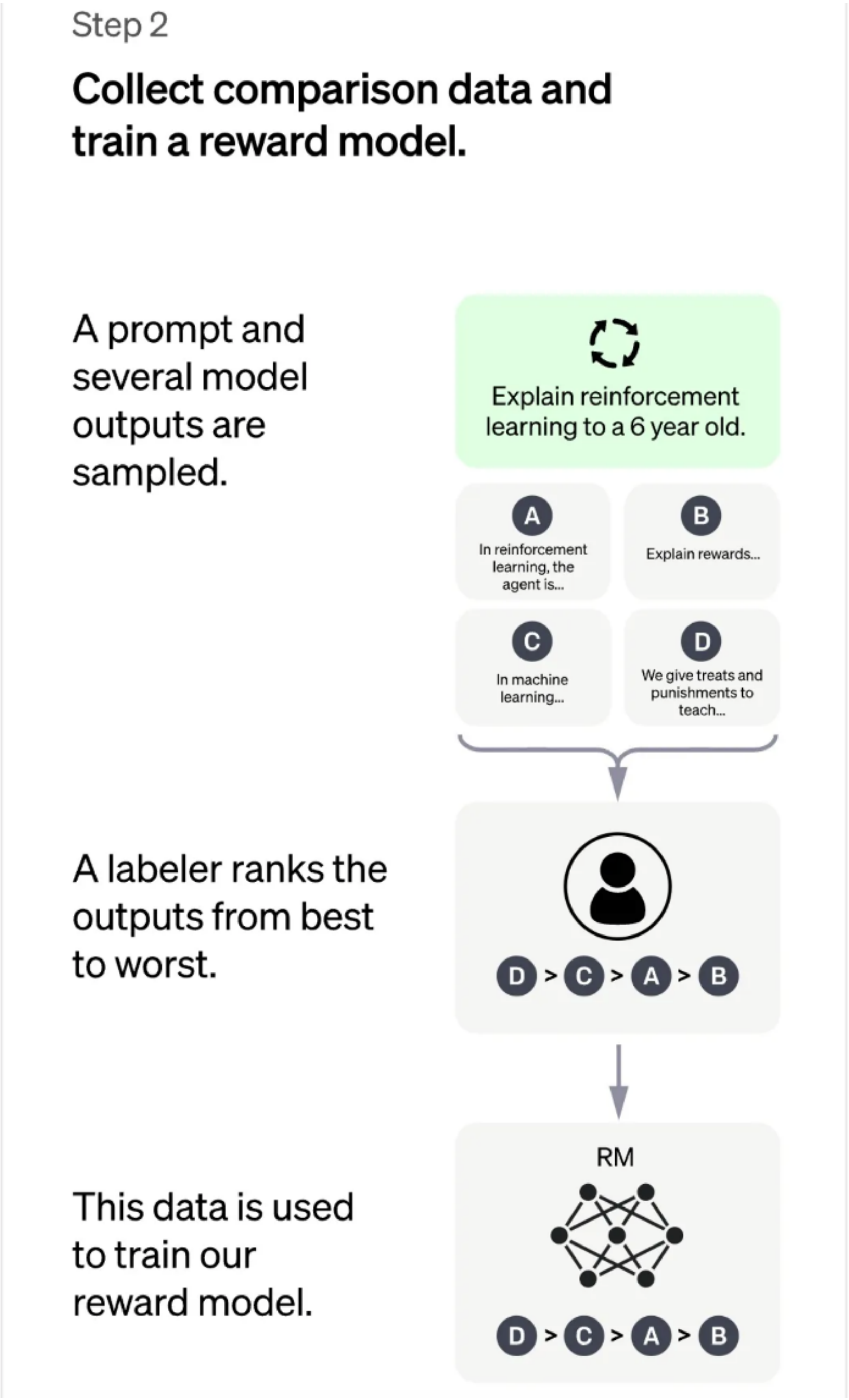
The policy generates an output.

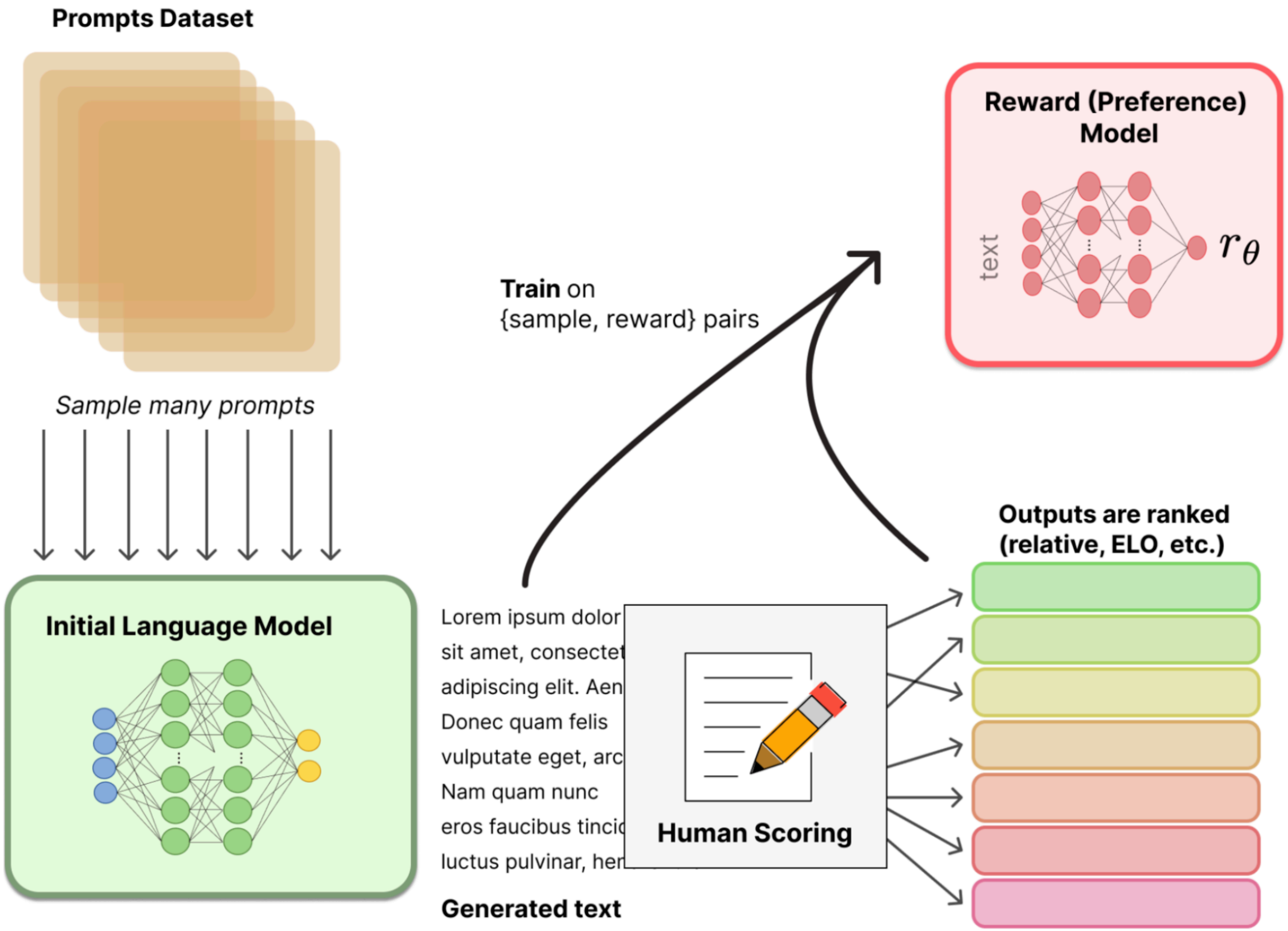
The reward model calculates a reward for the output.

The reward is used to update the policy using PPO.

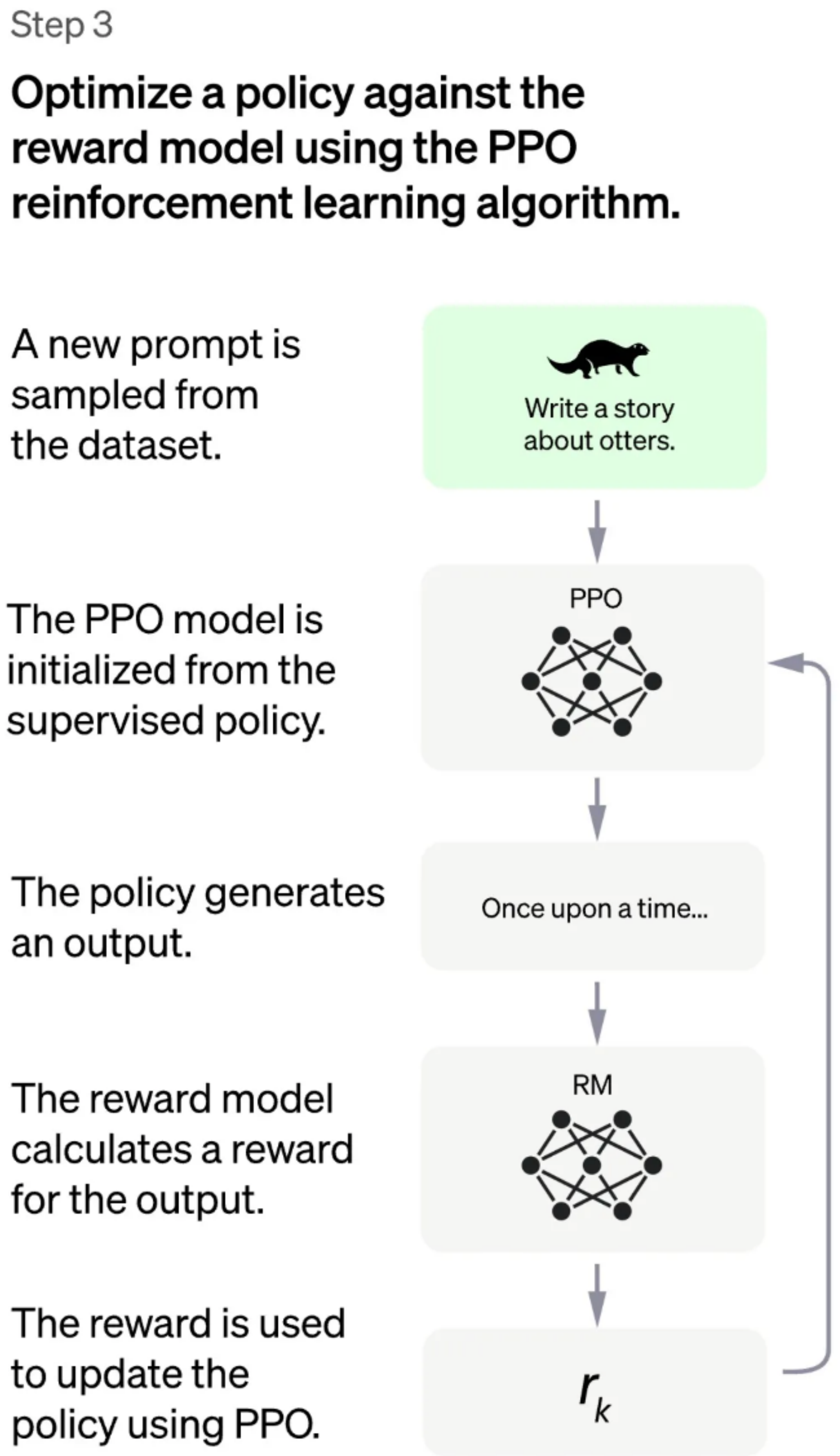


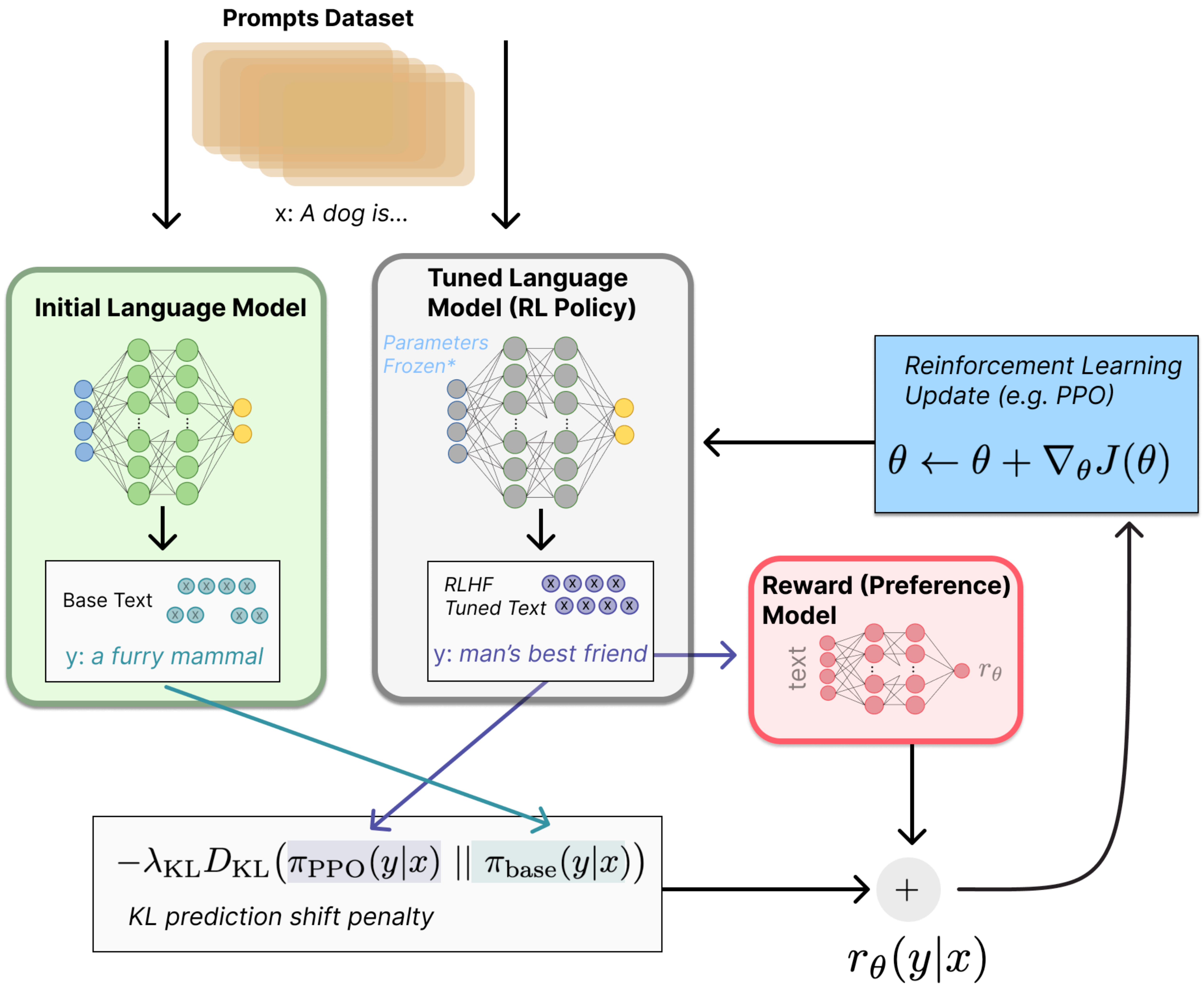
Reward Modeling





Reinforcement Learning





RLHF: Pros, Challenges and Variants

RLHF: Pros, Challenges and Variants

- RLHF works perhaps because it lets the model explore (Diversity hypothesis)

RLHF: Pros, Challenges and Variants

- RLHF works perhaps because it lets the model explore (Diversity hypothesis)
    - Not only training on ground truth

RLHF: Pros, Challenges and Variants

- RLHF works perhaps because it lets the model explore (Diversity hypothesis)
  - Not only training on ground truth
- Important step towards LLM safety

RLHF: Pros, Challenges and Variants

- RLHF works perhaps because it lets the model explore (Diversity hypothesis)
  - Not only training on ground truth
- Important step towards LLM safety
- Challenges:

RLHF: Pros, Challenges and Variants

- RLHF works perhaps because it lets the model explore (Diversity hypothesis)
  - Not only training on ground truth
- Important step towards LLM safety
- Challenges:
  - Data: Annotating preferences is tricky - Remember human eval?

RLHF: Pros, Challenges and Variants

- RLHF works perhaps because it lets the model explore (Diversity hypothesis)
  - Not only training on ground truth
- Important step towards LLM safety
- Challenges:
  - Data: Annotating preferences is tricky - Remember human eval?
  - Model: Hyperparameters are important

RLHF: Pros, Challenges and Variants

- RLHF works perhaps because it lets the model explore (Diversity hypothesis)
  - Not only training on ground truth
- Important step towards LLM safety
- Challenges:
  - Data: Annotating preferences is tricky - Remember human eval?
  - Model: Hyperparameters are important
  - Reward hacking: As model gets updated, the reward model may become incapable of scoring model outputs adequately

RLHF: Pros, Challenges and Variants

- RLHF works perhaps because it lets the model explore (Diversity hypothesis)
  - Not only training on ground truth
- Important step towards LLM safety
- Challenges:
  - Data: Annotating preferences is tricky - Remember human eval?
  - Model: Hyperparameters are important
  - Reward hacking: As model gets updated, the reward model may become incapable of scoring model outputs adequately
- As a result, there is a move towards non-RL methods for preference tuning

RLHF: Pros, Challenges and Variants

- RLHF works perhaps because it lets the model explore (Diversity hypothesis)
  - Not only training on ground truth
- Important step towards LLM safety
- Challenges:
  - Data: Annotating preferences is tricky - Remember human eval?
  - Model: Hyperparameters are important
  - Reward hacking: As model gets updated, the reward model may become incapable of scoring model outputs adequately
- As a result, there is a move towards non-RL methods for preference tuning
  - Note: Preference tuning is helpful, but whether RL is necessary for it remains an open question

Advanced Topics on LLMs

Lecture Outline

- ~~Quiz 6~~
- ~~Recap: Modern LLM Recipe~~
- ~~Recap: Alignment~~
- Advanced Topics (Highlights):
  - Pretraining data for LLMs
  - Evaluation of LLMs
    - LLM Harms
  - Beyond "Language" Models
    - Multimodal models
    - Multilingual models
    - LLMs + Retrievers
  - LLMs and Scaling Laws
  - LLMs as Mixtures of Experts

Training Data for LLMs

Training Data for LLMs

- Pre- and post-training

Training Data for LLMs

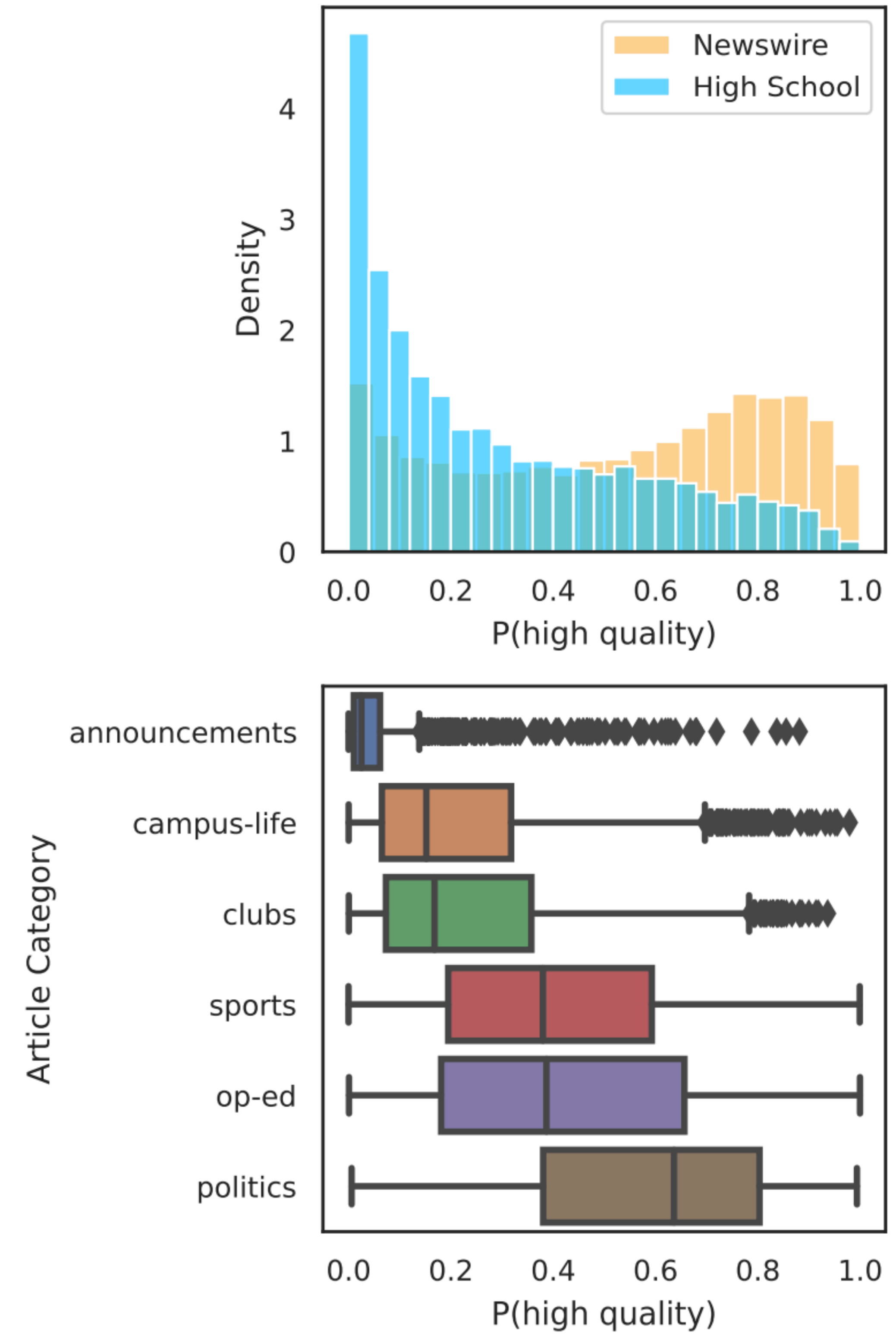
- Pre- and post-training
- Pretraining needs a lot of raw text

Training Data for LLMs

- Pre- and post-training
- Pretraining needs a lot of raw text
  - Crawled from the Web: Natural Solution

Training Data for LLMs

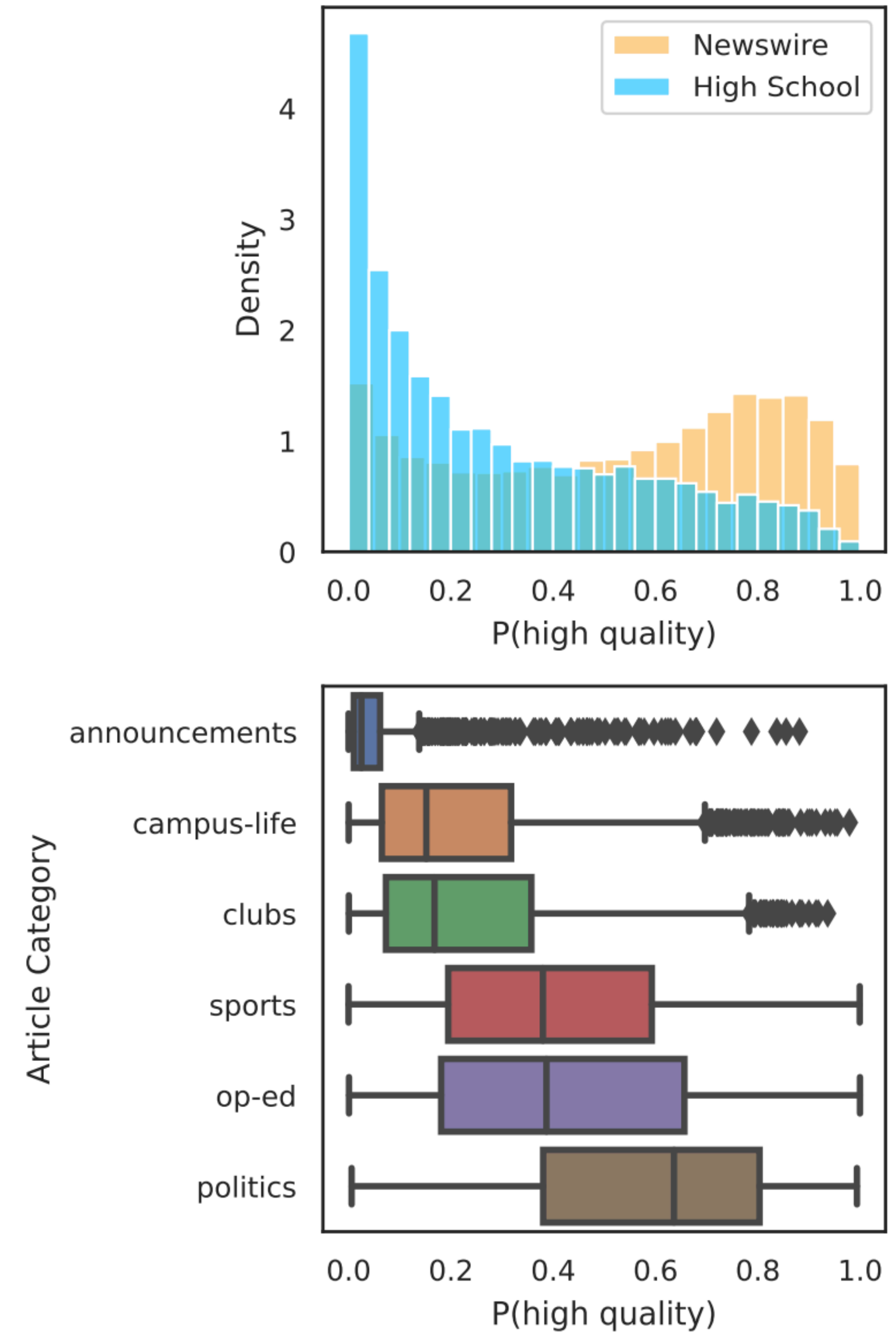
- Pre- and post-training
- Pretraining needs a lot of raw text
  - Crawled from the Web: Natural Solution
  - But not all data crawled from the web is good for LM training



Gururangan et al., 2022. https://arxiv.org/abs/2201.10474

Training Data for LLMs

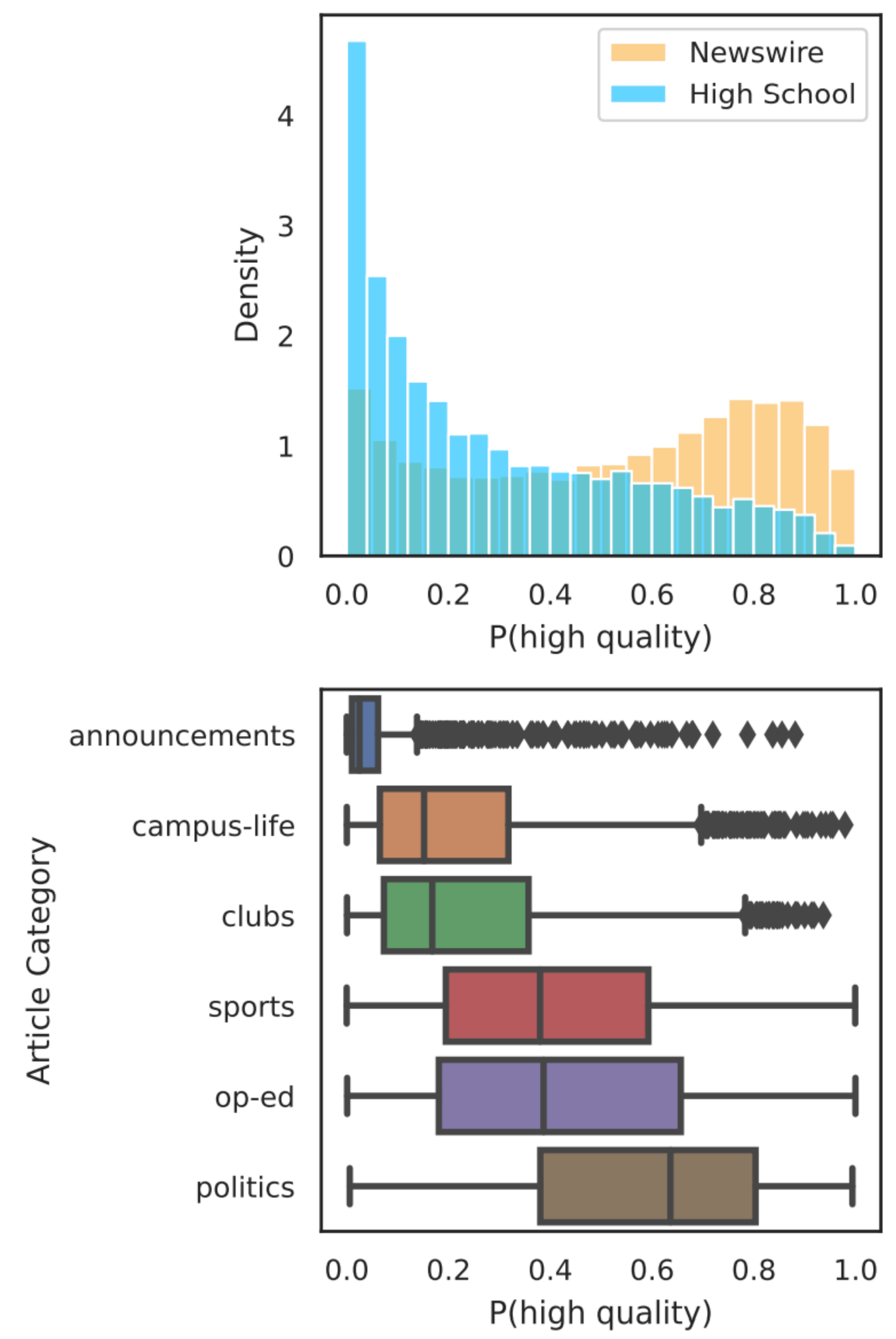
- Pre- and post-training
- Pretraining needs a lot of raw text
  - Crawled from the Web: Natural Solution
  - But not all data crawled from the web is good for LM training
  - Quality Filters!



Gururangan et al., 2022. https://arxiv.org/abs/2201.10474

Training Data for LLMs

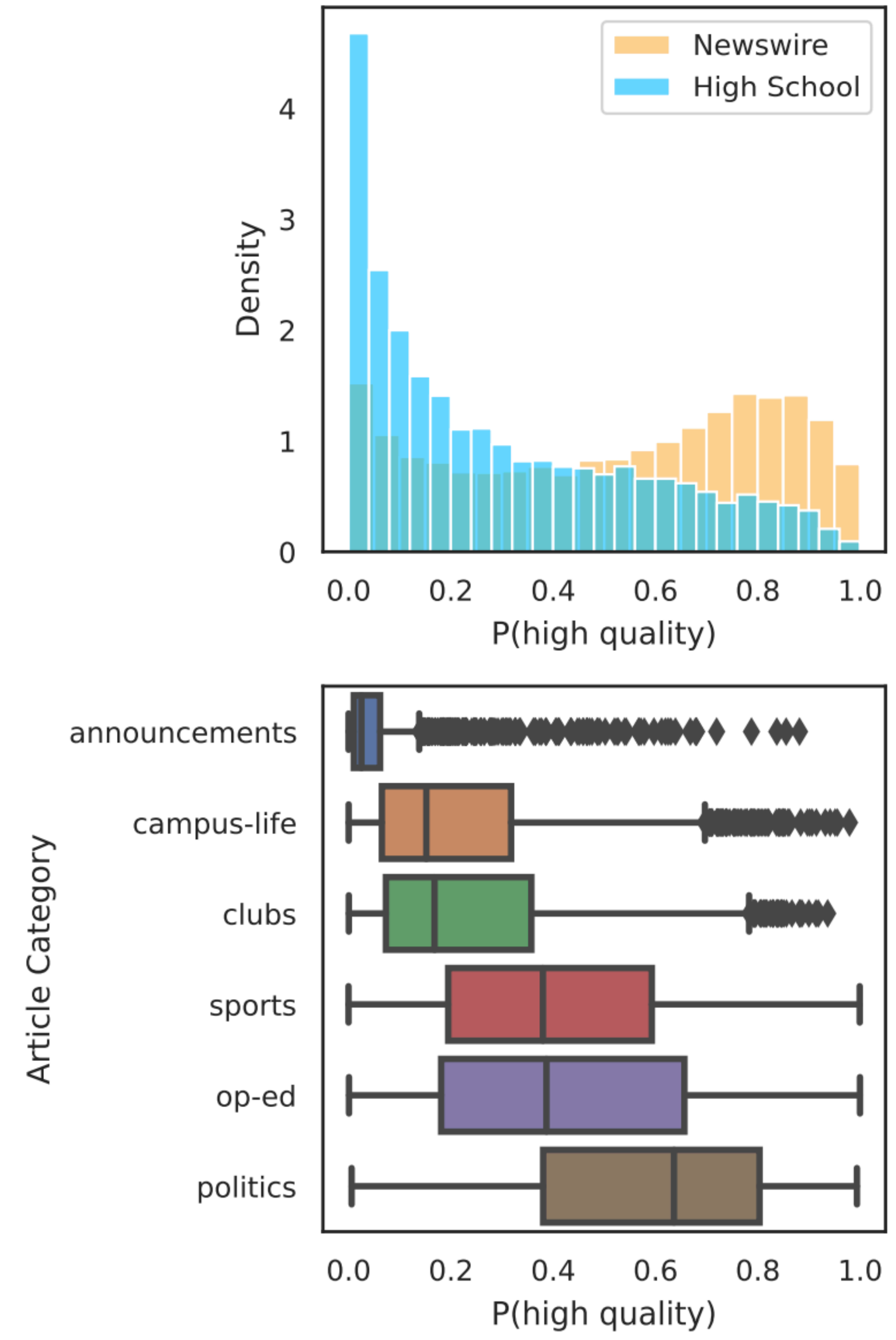
- Pre- and post-training
- Pretraining needs a lot of raw text
  - Crawled from the Web: Natural Solution
  - But not all data crawled from the web is good for LM training
  - Quality Filters!
- Post-training



Gururangan et al., 2022. https://arxiv.org/abs/2201.10474

Training Data for LLMs

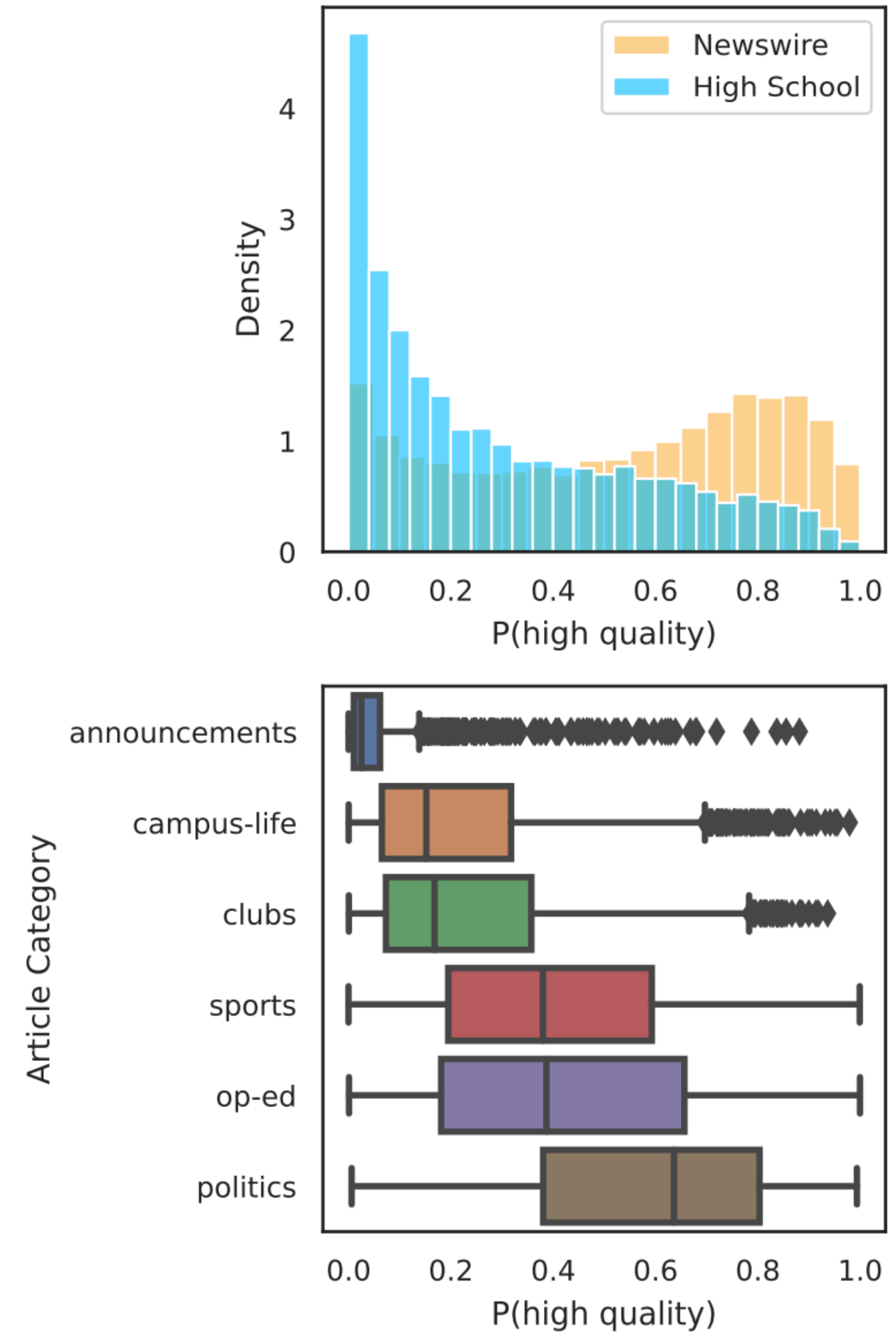
- Pre- and post-training
- Pretraining needs a lot of raw text
  - Crawled from the Web: Natural Solution
  - But not all data crawled from the web is good for LM training
  - Quality Filters!
- Post-training
  - Instruction Tuning Data



Gururangan et al., 2022. https://arxiv.org/abs/2201.10474

Training Data for LLMs

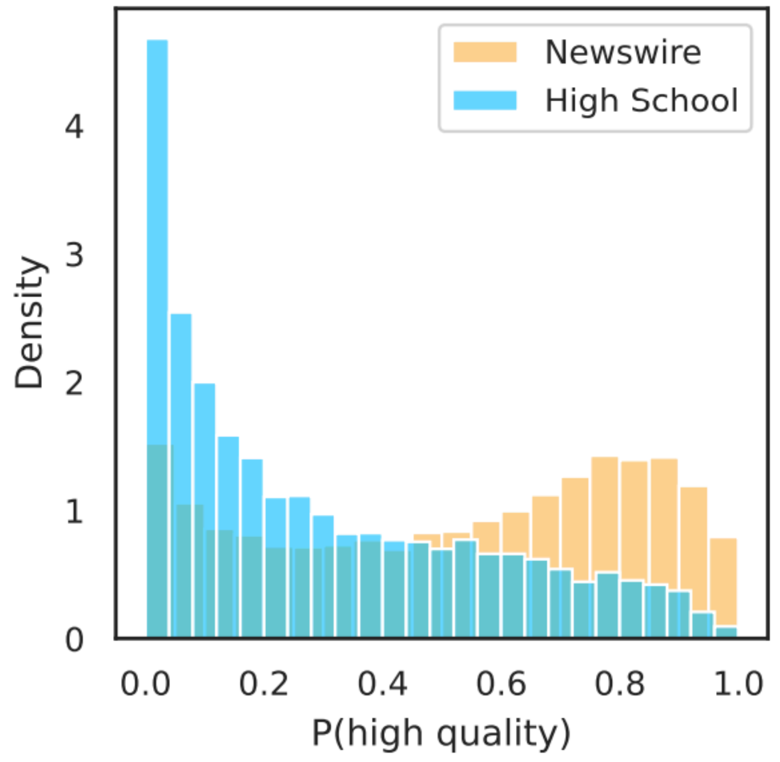
- Pre- and post-training
- Pretraining needs a lot of raw text
  - Crawled from the Web: Natural Solution
  - But not all data crawled from the web is good for LM training
  - Quality Filters!
- Post-training
  - Instruction Tuning Data
  - Preference Data for RLHF

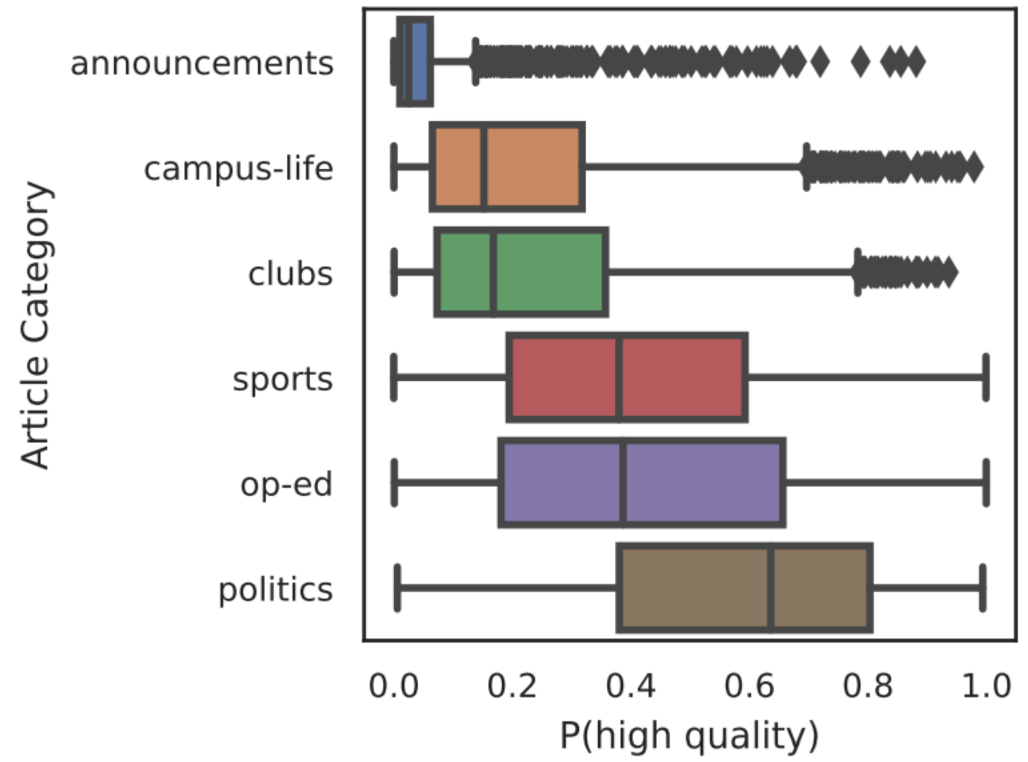


Gururangan et al., 2022. https://arxiv.org/abs/2201.10474

Training Data for LLMs

- Pre- and post-training
- Pretraining needs a lot of raw text
  - Crawled from the Web: Natural Solution
  - But not all data crawled from the web is good for LM training
  - Quality Filters!
- Post-training
  - Instruction Tuning Data
  - Preference Data for RLHF
  - Not easy to produce…





Gururangan et al., 2022. https://arxiv.org/abs/2201.10474

Pretraining Data

Pretraining Data

- Language models are trained on “raw text”

Pretraining Data

- Language models are trained on “raw text”
- To be highly capable (e.g., have linguistic and world knowledge), this text should span a **broad** range of domains, genres, languages, etc.

Pretraining Data

- Language models are trained on "raw text"
- To be highly capable (e.g., have linguistic and world knowledge), this text should span a **broad** range of domains, genres, languages, etc.
- A natural place (but not the only place) to look for such text is the **web**
    - Google search index is 100 petabytes; the actual web is likely even larger
    - **Private datasets** owned by big companies are even larger! WalMart generates 2.5 petabytes of data each hour!

Pretraining Data

- Language models are trained on “raw text”
- To be highly capable (e.g., have linguistic and world knowledge), this text should span a **broad** range of domains, genres, languages, etc.
- A natural place (but not the only place) to look for such text is the **web**
  - Google search index is 100 petabytes; the actual web is likely even larger
  - **Private datasets** owned by big companies are even larger! WalMart generates 2.5 petabytes of data each hour!

- **Common Crawl** is a nonprofit organization that crawls the web and provides snapshots that are free to the public
  - Standard source of data to train many models such as T5, GPT-3, etc.
  - The April 2021 snapshot of Common Crawl has 320 TB

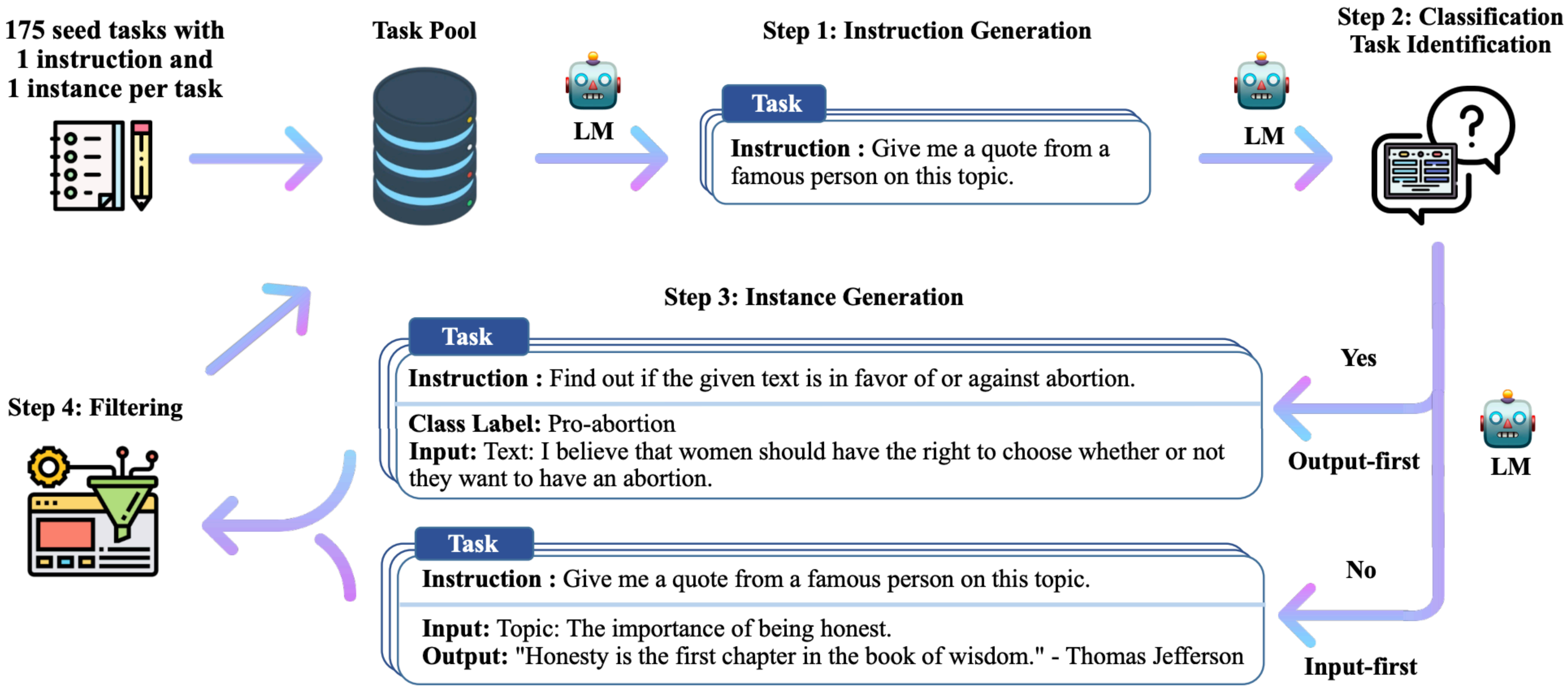
COMMON CRAWL

Pretraining Data

- Language models are trained on “raw text”
- To be highly capable (e.g., have linguistic and world knowledge), this text should span a **broad** range of domains, genres, languages, etc.
- A natural place (but not the only place) to look for such text is the **web**
  - Google search index is 100 petabytes; the actual web is likely even larger
  - **Private datasets** owned by big companies are even larger! WalMart generates 2.5 petabytes of data each hour!

- **Common Crawl** is a nonprofit organization that crawls the web and provides snapshots that are free to the public
  - Standard source of data to train many models such as T5, GPT-3, etc.
  - The April 2021 snapshot of Common Crawl has 320 TB
- The Colossal Clean Crawled Corpus (C4) is a larger was created to train the T5 model — 806 GB / 156 billion tokens

Instruction Tuning Data



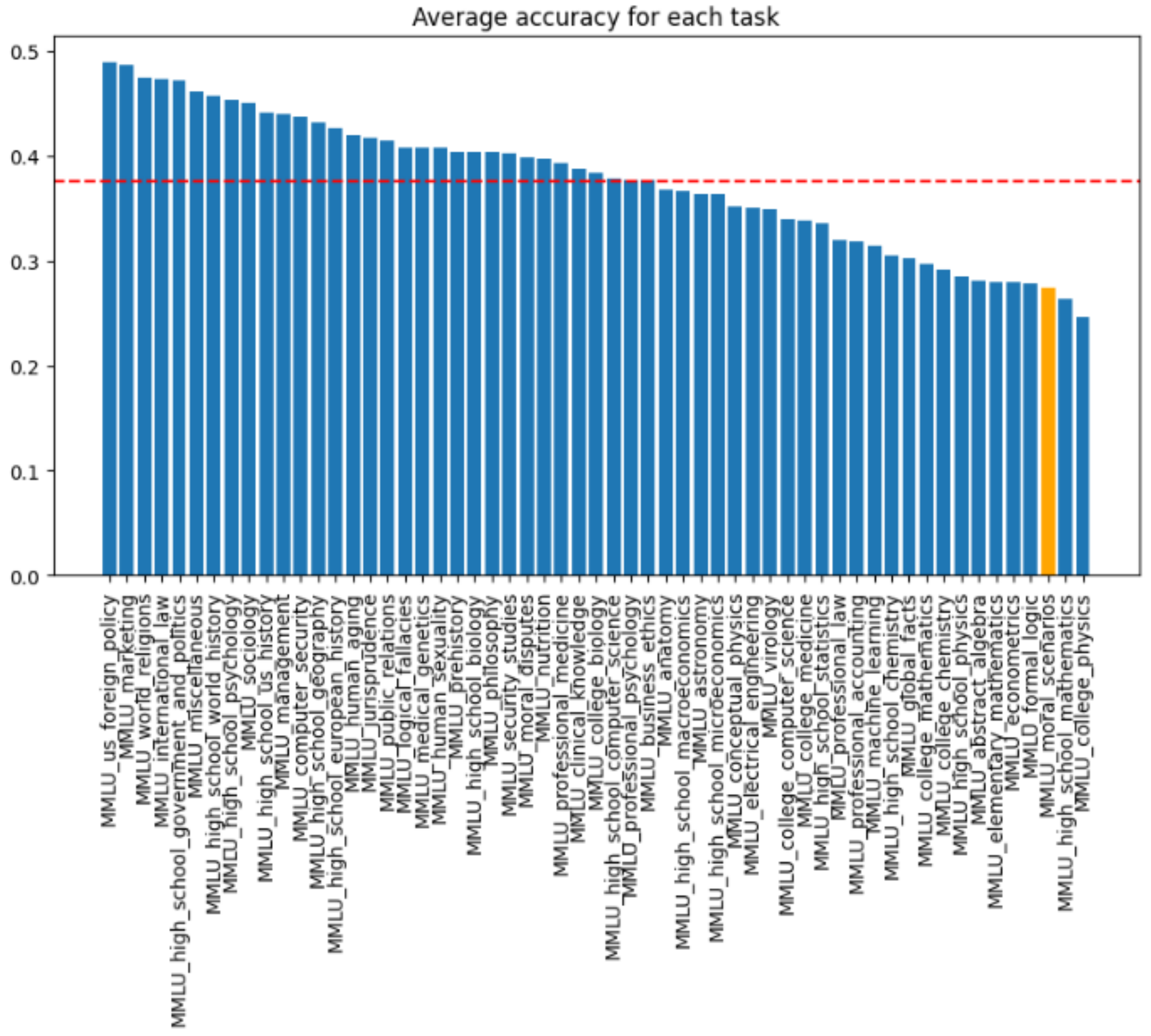
Wang et al., 2022. https://arxiv.org/abs/2212.10560

Evaluation of LLMs

- Almost exclusively on downstream tasks, as opposed to intrinsic metrics
  - Intrinsic metrics, e.g. perplexity
- 4-5 Multitask benchmarks
  - GLUE - Language Understanding Tasks
  - SuperGLUE - Language Understanding Tasks
  - HellaSwag - Commonsense Reasoning
  - Truthful QA - Fact Verification
  - MMLU

Evaluation of LLMs

- Almost exclusively on downstream tasks, as opposed to intrinsic metrics
  - Intrinsic metrics, e.g. perplexity
- 4-5 Multitask benchmarks
  - GLUE - Language Understanding Tasks
  - SuperGLUE - Language Understanding Tasks
  - HellaSwag - Commonsense Reasoning
  - Truthful QA - Fact Verification
  - MMLU



LLMs: Harms

LLMs: Harms

- 3 main categories of harms:

LLMs: Harms

- 3 main categories of harms:
    - Hallucinations: LLM generations may contain falsehoods

The ChatGPT Lawyer Explains Himself

In a cringe-inducing court hearing, a lawyer who relied on A.I. to craft a motion full of made-up case law said he "did not comprehend" that the chat bot could lead him astray.

Steven A. Schwartz told a judge considering sanctions that the episode had been "deeply embarrassing." Jefferson Siegel for The New York Times

LLMs: Harms

- 3 main categories of harms:
  - Hallucinations: LLM generations may contain falsehoods
  - Data Privacy and Copyright Violations

The ChatGPT Lawyer Explains Himself

In a cringe-inducing court hearing, a lawyer who relied on A.I. to craft a motion full of made-up case law said he "did not comprehend" that the chat bot could lead him astray.

Steven A. Schwartz told a judge considering sanctions that the episode had been "deeply embarrassing." Jefferson Siegel for The New York Times

LLMs: Harms

- 3 main categories of harms:
    - Hallucinations: LLM generations may contain falsehoods
    - Data Privacy and Copyright Violations
    - Social Biases (a lot comes from the data!)

The ChatGPT Lawyer Explains Himself

In a cringe-inducing court hearing, a lawyer who relied on A.I. to craft a motion full of made-up case law said he "did not comprehend" that the chat bot could lead him astray.

Steven A. Schwartz told a judge considering sanctions that the episode had been "deeply embarrassing." Jefferson Siegel for The New York Times

'He Would Still Be Here': Man Dies by Suicide After Talking with AI Chatbot, Widow Says

LLMs: Harms

- 3 main categories of harms:
  - Hallucinations: LLM generations may contain falsehoods
  - Data Privacy and Copyright Violations
  - Social Biases (a lot comes from the data!)
- Others: Access due to high costs; Interpretability

The ChatGPT Lawyer Explains Himself

In a cringe-inducing court hearing, a lawyer who relied on A.I. to craft a motion full of made-up case law said he "did not comprehend" that the chat bot could lead him astray.

Steven A. Schwartz told a judge considering sanctions that the episode had been "deeply embarrassing." Jefferson Siegel for The New York Times

VICE

'He Would Still Be Here': Man Dies by Suicide After Talking with AI Chatbot, Widow Says

LLMs: Harms

- 3 main categories of harms:
  - Hallucinations: LLM generations may contain falsehoods
  - Data Privacy and Copyright Violations
  - Social Biases (a lot comes from the data!)
- Others: Access due to high costs; Interpretability
- Harms arising due to LLM behaviors

The ChatGPT Lawyer Explains Himself

In a cringe-inducing court hearing, a lawyer who relied on A.I. to craft a motion full of made-up case law said he "did not comprehend" that the chat bot could lead him astray.

Steven A. Schwartz told a judge considering sanctions that the episode had been "deeply embarrassing." Jefferson Siegel for The New York Times

VICE

'He Would Still Be Here': Man Dies by Suicide After Talking with AI Chatbot, Widow Says

LLMs: Harms

- 3 main categories of harms:
    - Hallucinations: LLM generations may contain falsehoods
    - Data Privacy and Copyright Violations
    - Social Biases (a lot comes from the data!)
- Others: Access due to high costs; Interpretability
- Harms arising due to LLM behaviors
    - Toxicity, Disinformation, Misinformation

The ChatGPT Lawyer Explains Himself

In a cringe-inducing court hearing, a lawyer who relied on A.I. to craft a motion full of made-up case law said he "did not comprehend" that the chat bot could lead him astray.

Steven A. Schwartz told a judge considering sanctions that the episode had been "deeply embarrassing." Jefferson Siegel for The New York Times

'He Would Still Be Here': Man Dies by Suicide After Talking with AI Chatbot, Widow Says

LLMs: Harms

- 3 main categories of harms:
    - Hallucinations: LLM generations may contain falsehoods
    - Data Privacy and Copyright Violations
    - Social Biases (a lot comes from the data!)
- Others: Access due to high costs; Interpretability
- Harms arising due to LLM behaviors
    - Toxicity, Disinformation, Misinformation
    - Representational and Allocational

The ChatGPT Lawyer Explains Himself

In a cringe-inducing court hearing, a lawyer who relied on A.I. to craft a motion full of made-up case law said he "did not comprehend" that the chat bot could lead him astray.

Steven A. Schwartz told a judge considering sanctions that the episode had been "deeply embarrassing." Jefferson Siegel for The New York Times

VICE

'He Would Still Be Here': Man Dies by Suicide After Talking with AI Chatbot, Widow Says

LLMs: Harms

- 3 main categories of harms:
  - Hallucinations: LLM generations may contain falsehoods
  - Data Privacy and Copyright Violations
  - Social Biases (a lot comes from the data!)
- Others: Access due to high costs; Interpretability
- Harms arising due to LLM behaviors
  - Toxicity, Disinformation, Misinformation
  - Representational and Allocational
- Dual Use

The ChatGPT Lawyer Explains Himself

In a cringe-inducing court hearing, a lawyer who relied on A.I. to craft a motion full of made-up case law said he "did not comprehend" that the chat bot could lead him astray.

Steven A. Schwartz told a judge considering sanctions that the episode had been "deeply embarrassing." Jefferson Siegel for The New York Times

VICE

'He Would Still Be Here': Man Dies by Suicide After Talking with AI Chatbot, Widow Says

Multimodal Models

Radford et al., 2021. https://arxiv.org/abs/2103.00020

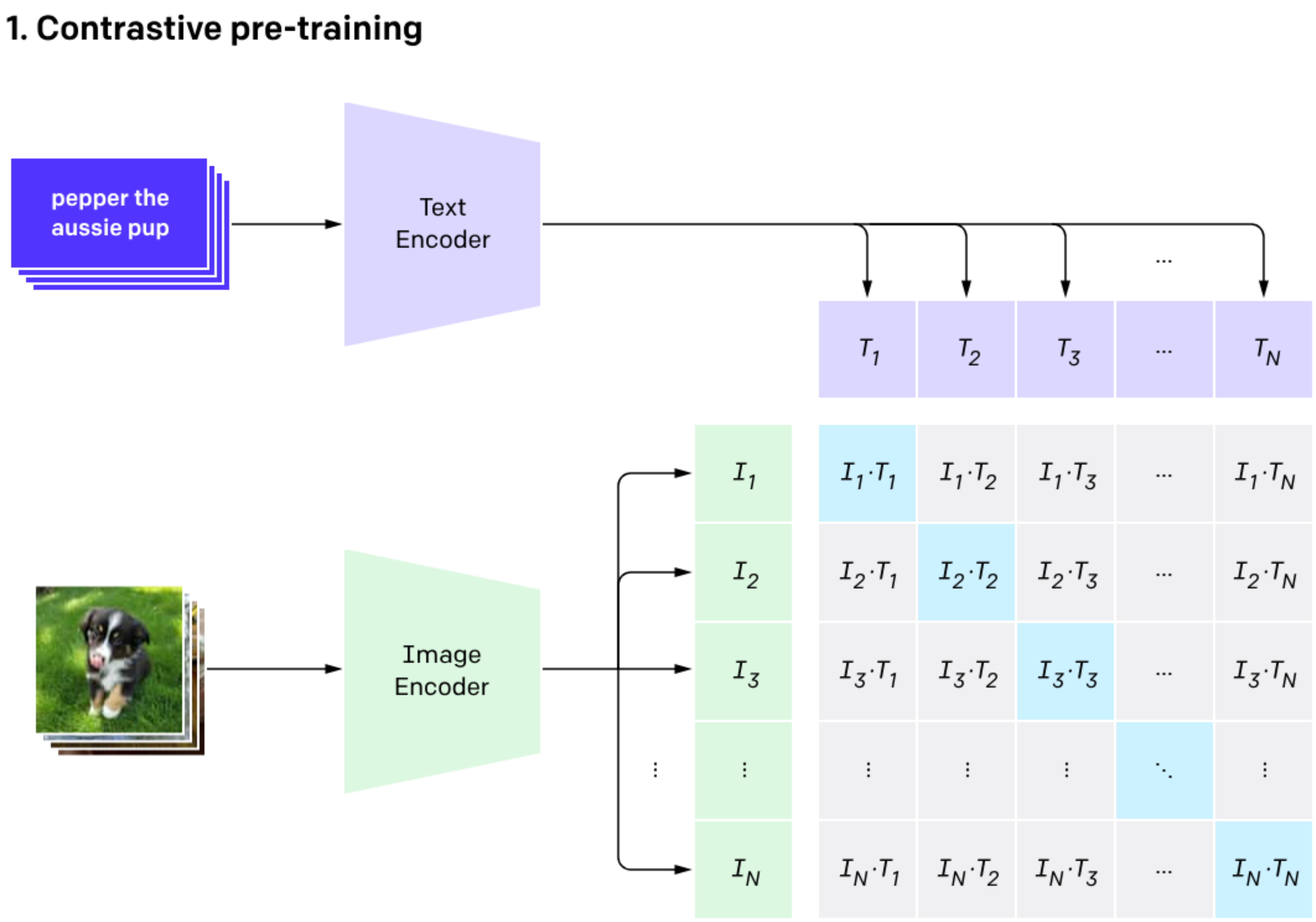
Multimodal Models

- Language + other modalities

Radford et al., 2021. https://arxiv.org/abs/2103.00020

Multimodal Models

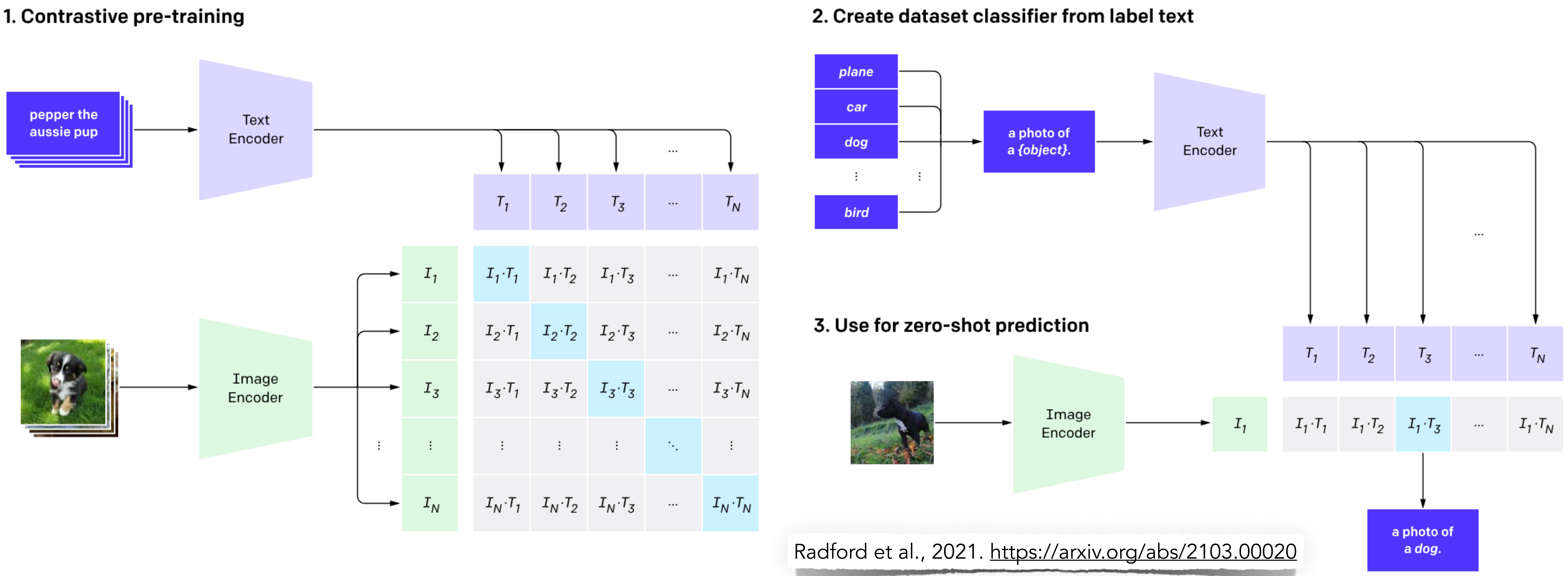
- Language + other modalities
  - e.g. CLIP from OpenAI (image recognition)



Radford et al., 2021. https://arxiv.org/abs/2103.00020

Multimodal Models

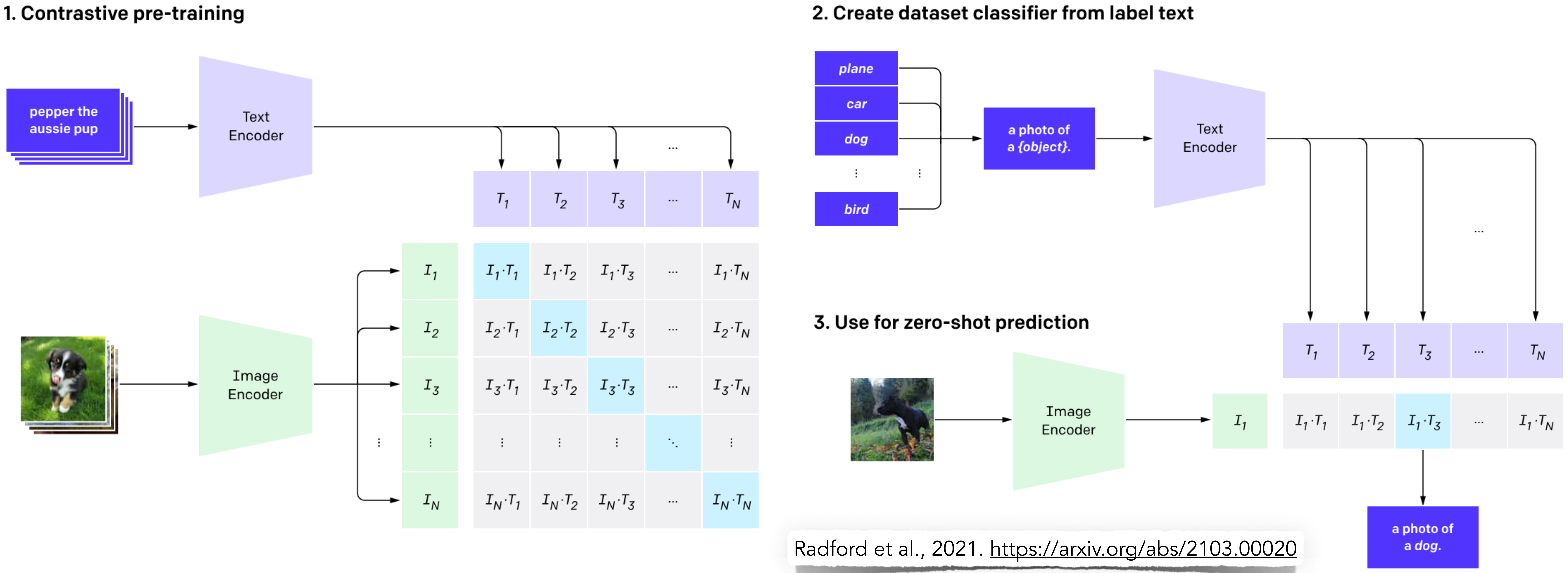
- Language + other modalities
  - e.g. CLIP from OpenAI (image recognition)



Radford et al., 2021. https://arxiv.org/abs/2103.00020

Multimodal Models

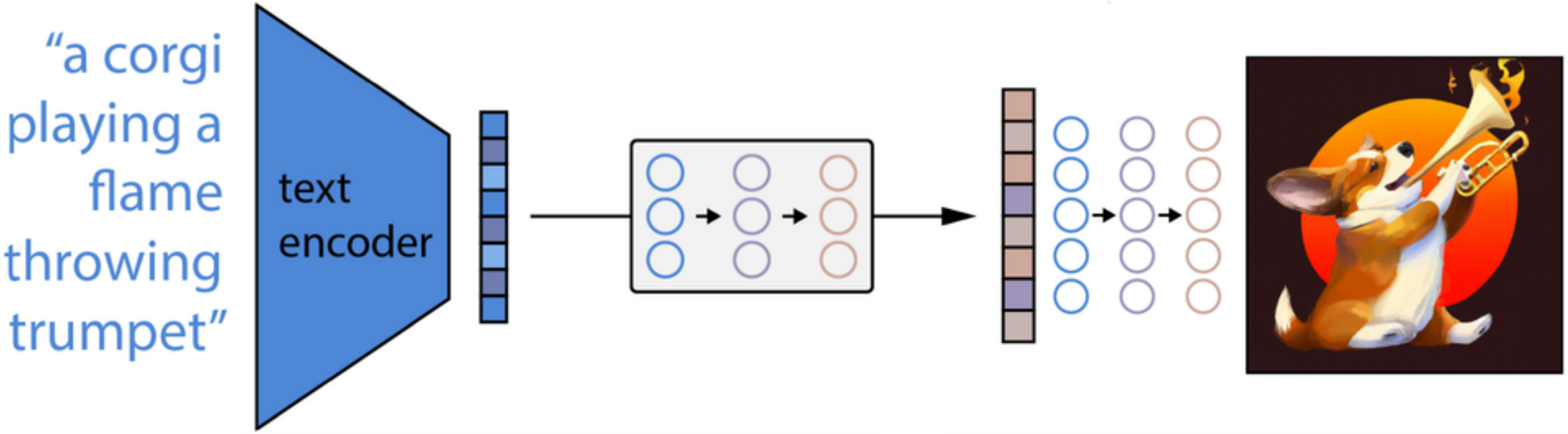
- Language + other modalities
  - e.g. CLIP from OpenAI (image recognition)
- Not just vision: audio, video, etc.



Radford et al., 2021. https://arxiv.org/abs/2103.00020

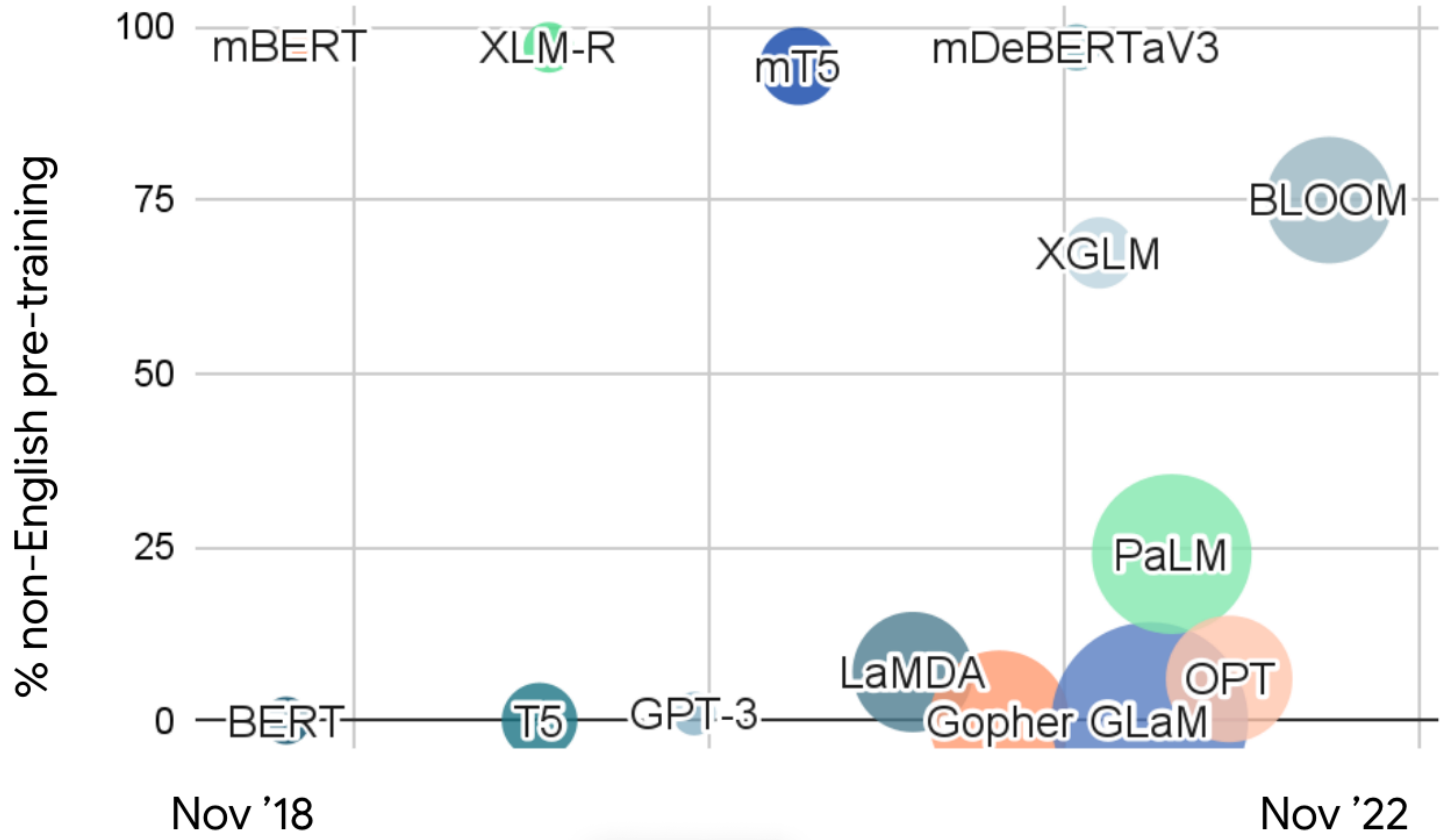
Multimodal models: DALL-E

- DALL-E 2 creates images and art from a description in natural language
- Text encoder, image decoder
- Uses CLIP to find text-image pairings in a high dimensional space



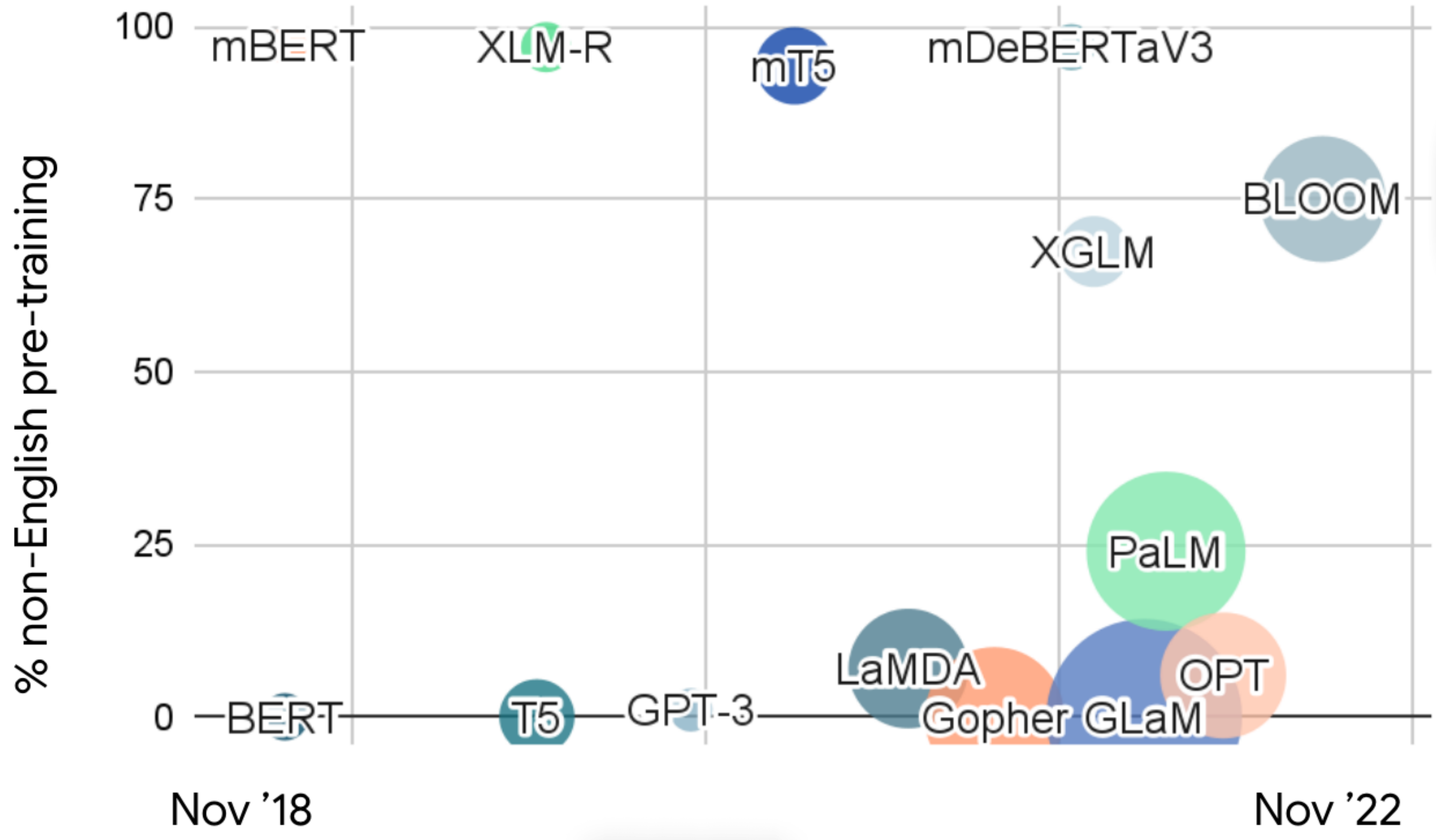
Ramesh et al., 2022. https://arxiv.org/abs/2204.06125 Source: https://www.assemblyai.com/blog/how-dall-e-2-actually-works/

Multilingual Models



Source: ruder.io

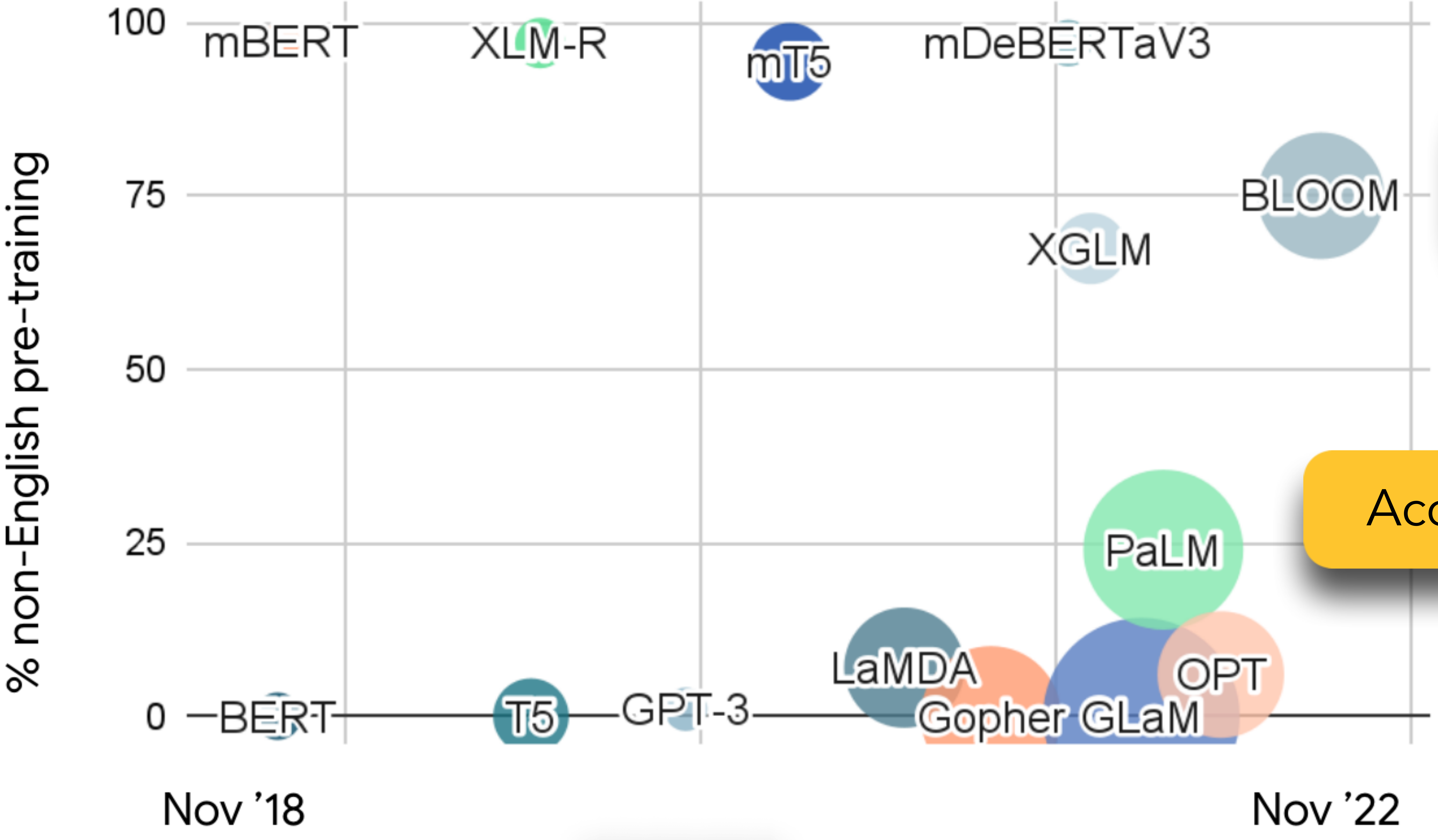
Multilingual Models



Source: ruder.io

Primary Challenge:
Availability of Data!

Multilingual Models



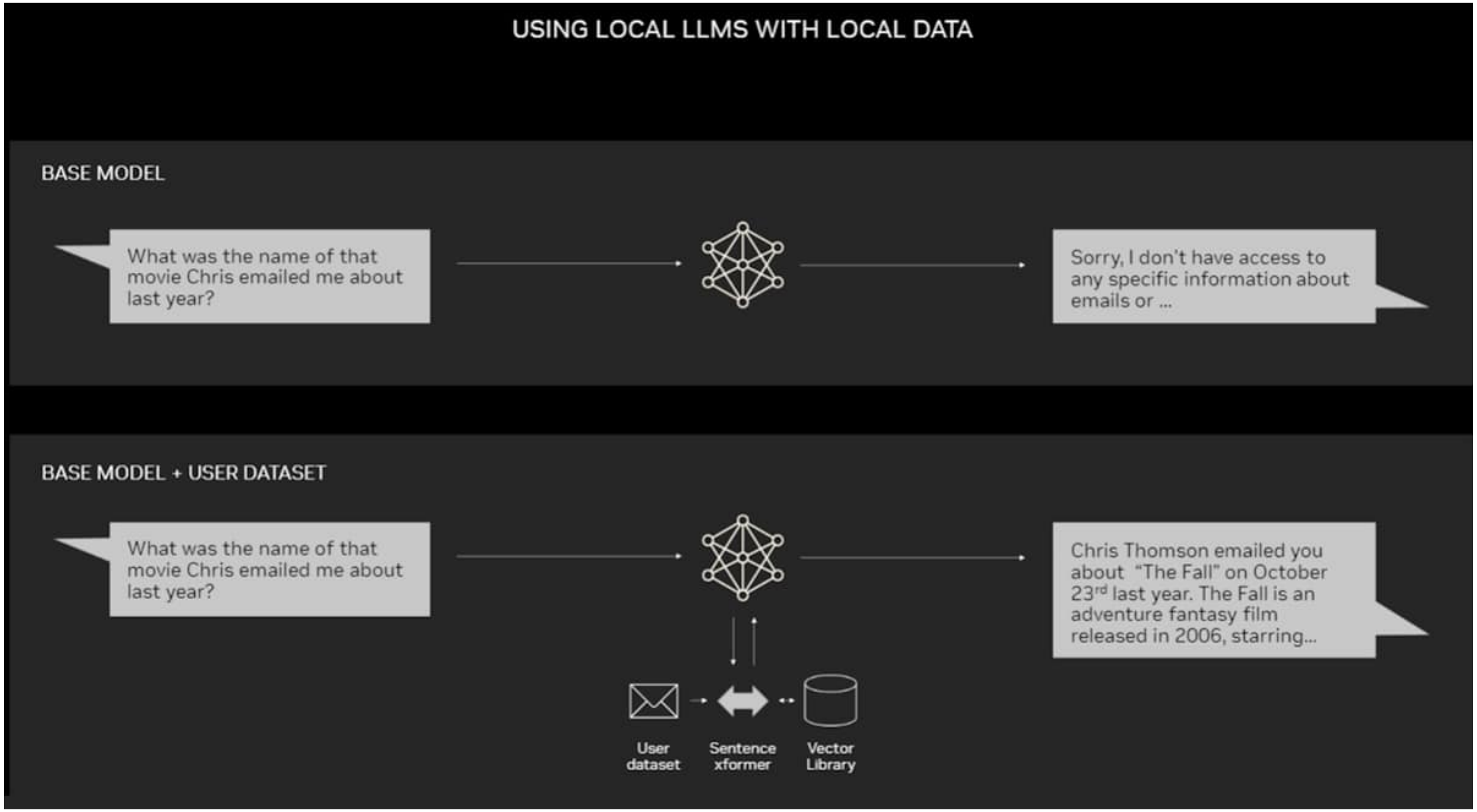
Source: ruder.io

Primary Challenge: Availability of Data!

Accidental Multilinguality

Retrieval + Generation

- Allows for a user-specified context through retrieval from a data store



RAG: Lewis et al., 2020 https://arxiv.org/pdf/2005.11401.pdf

Source: https://blogs.nvidia.com/blog/what-is-retrieval-augmented-generation/

USC Viterbi

Context Lengths

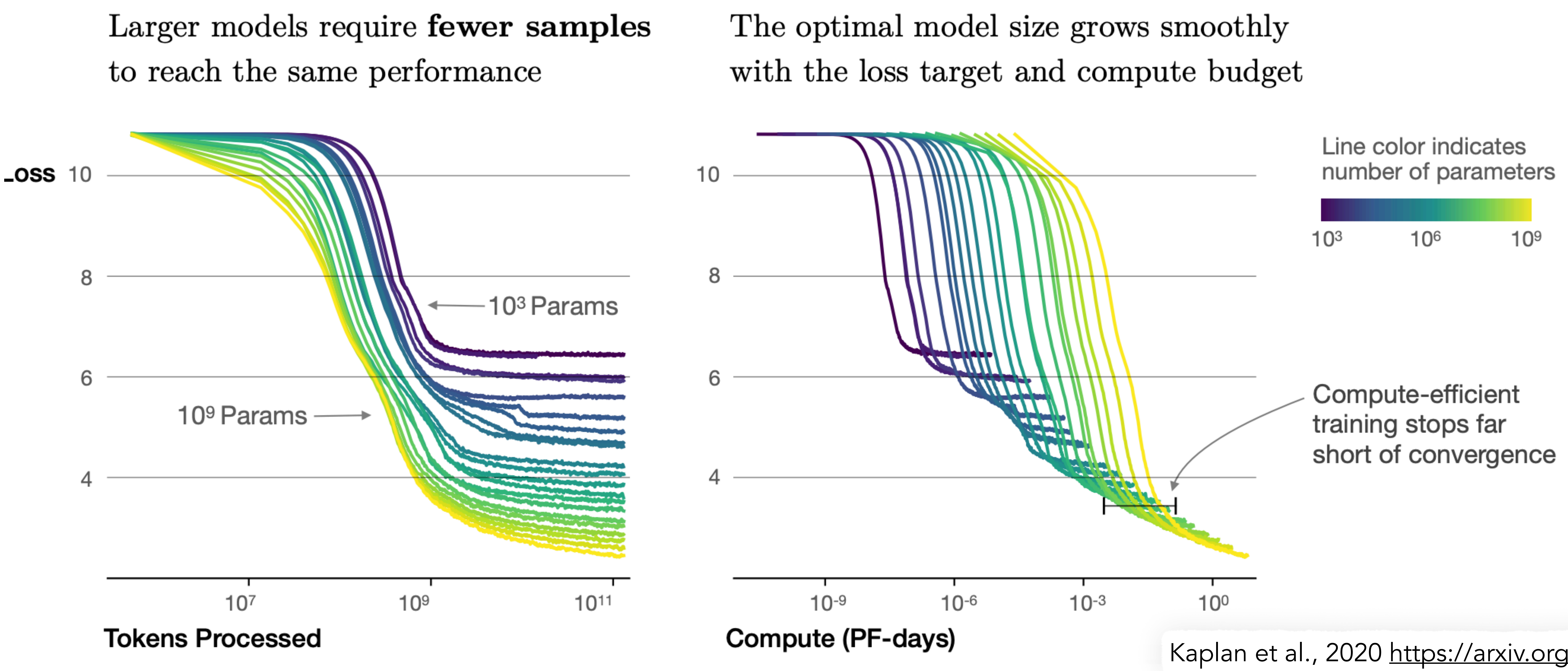
- GPT-2 has a context length of 1024 tokens

OpenAI model's version	GPT-3 (ada, babbage, curie, davinci)	GPT-3.5 (gpt-3.5-turbo, gpt-3.5-turbo-0301,text-davinci-003, text-davinci-002)*	GPT-4-8K
Context length (max request)	2,049	4,096	8,192
Number of English words	~1,500	~3,000	~6,000
Number of single-spaced pages of English text	3	6	12

Source: Neoteric

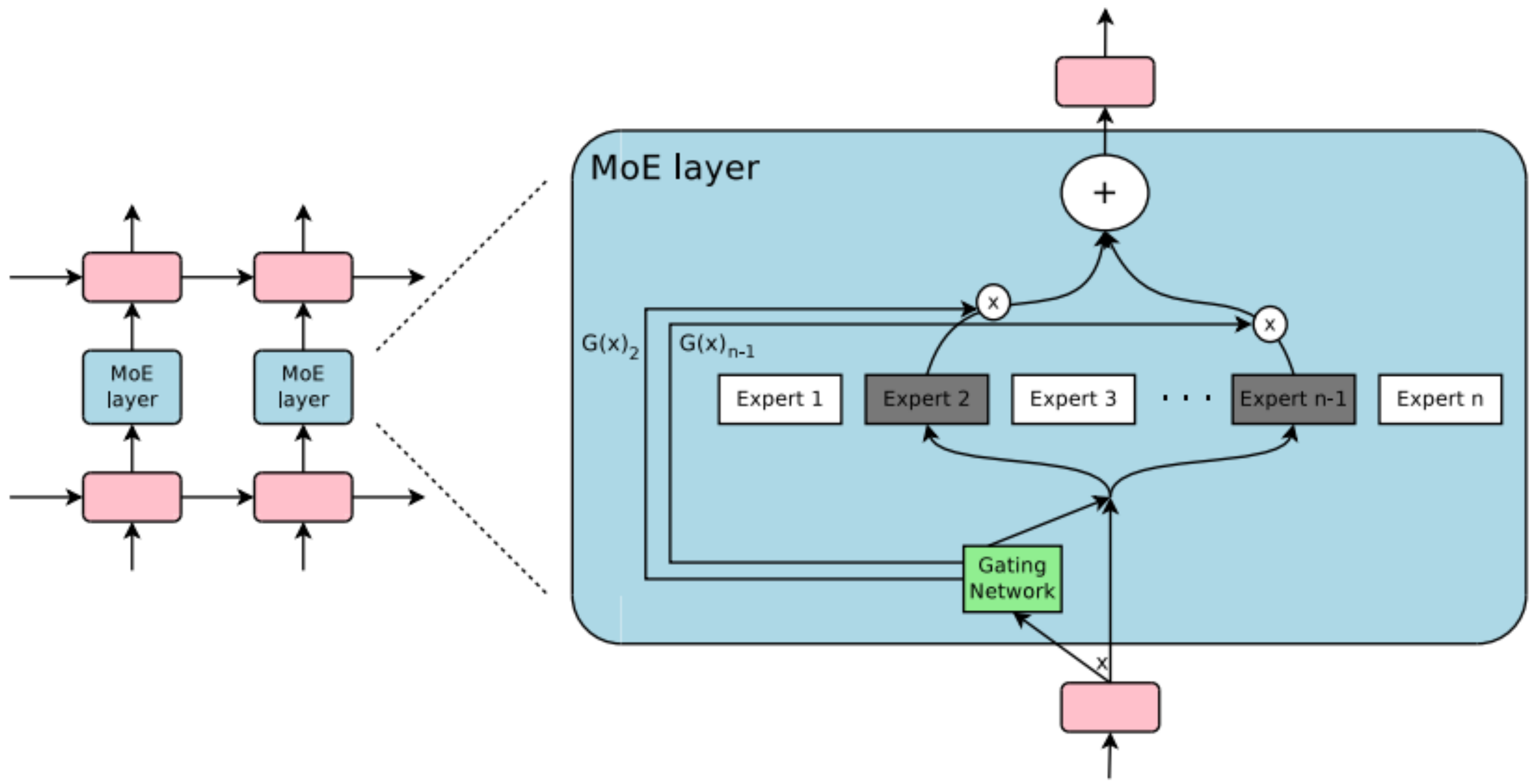
LLMs + Scaling Laws

- Predictive rules of model performance, given parameter size, data size, etc.



Kaplan et al., 2020 https://arxiv.org/abs/2001.08361

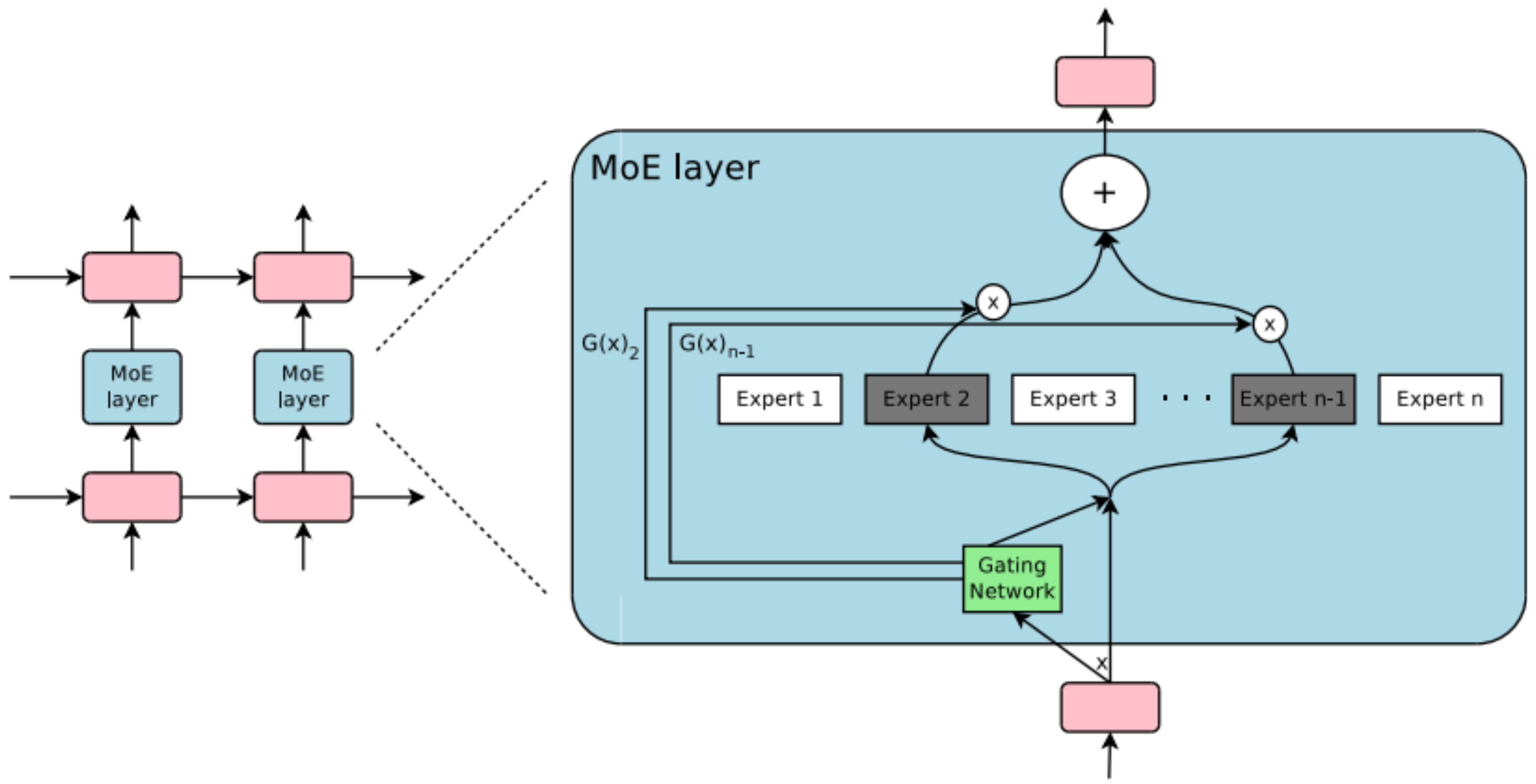
LLMs as Mixtures of Experts



Source: Huggingface

USC Viterbi

LLMs as Mixtures of Experts



Source: Huggingface

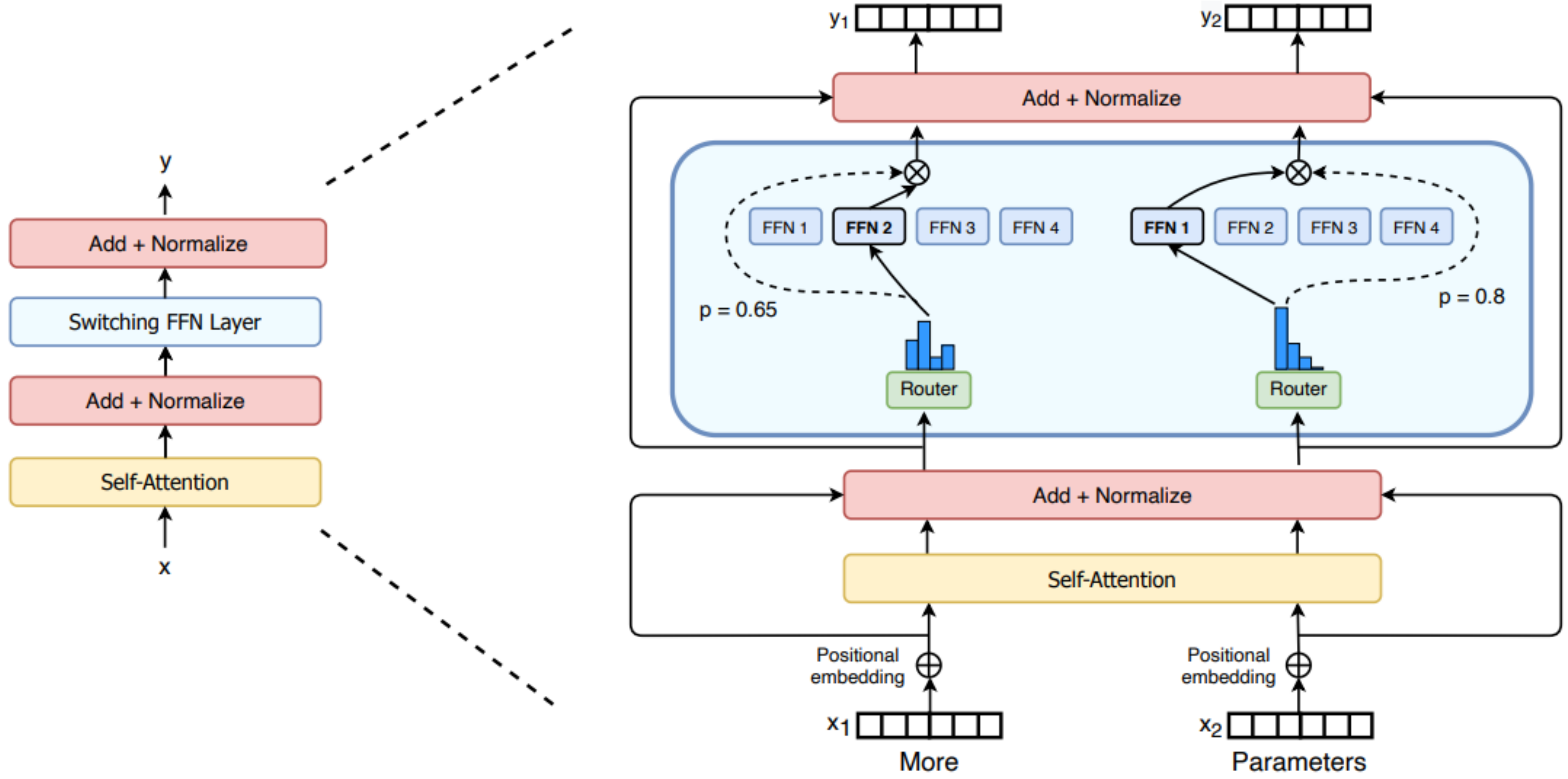


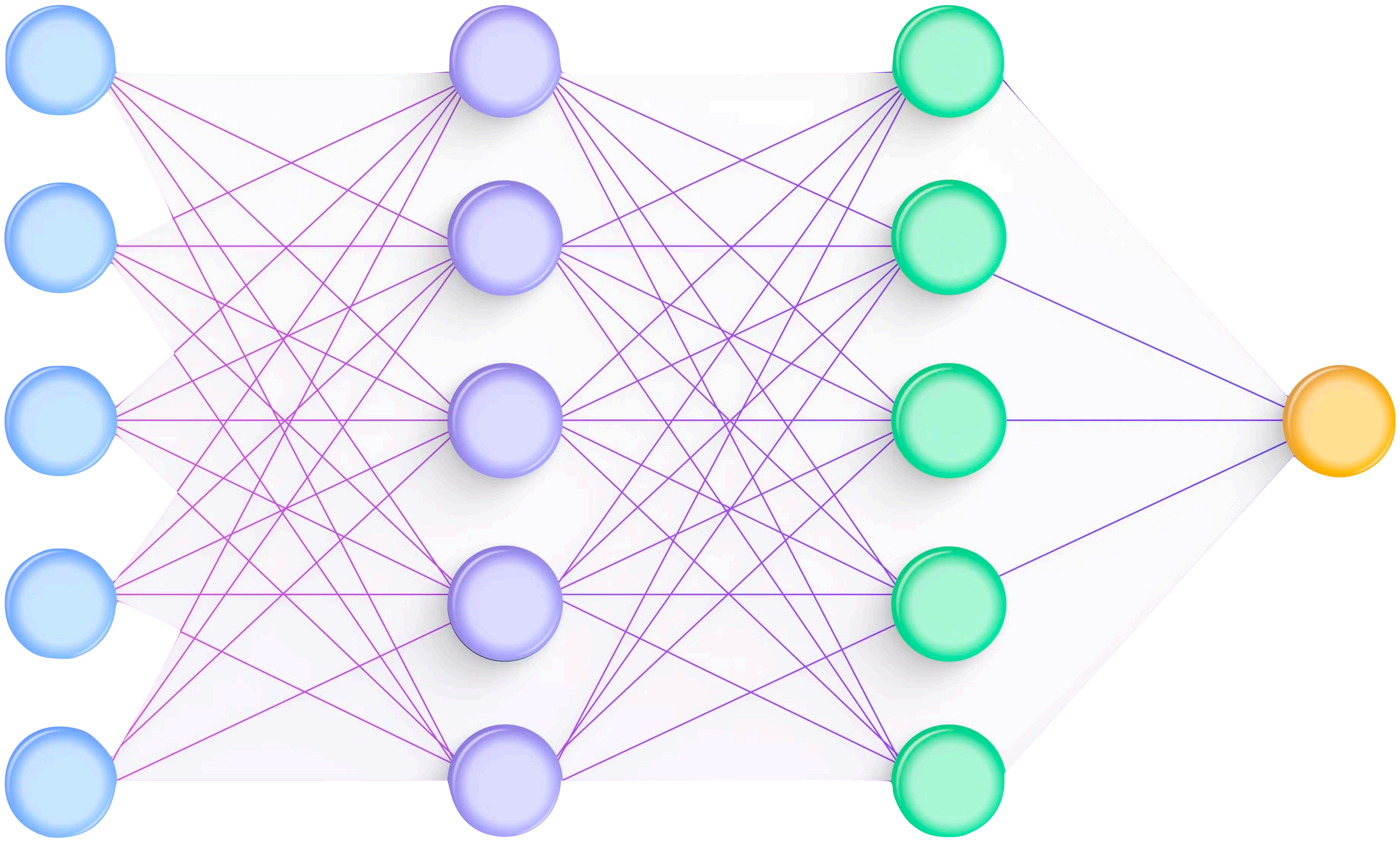
Figure 2: Illustration of a Switch Transformer encoder block. We replace the dense feed forward network (FFN) layer present in the Transformer with a sparse Switch FFN layer (light blue). The layer operates independently on the tokens in the sequence. We diagram two tokens (x_1 = "More" and x_2 = "Parameters" below) being routed (solid lines) across four FFN experts, where the router independently routes each token. The switch FFN layer returns the output of the selected FFN multiplied by the router gate value (dotted-line).

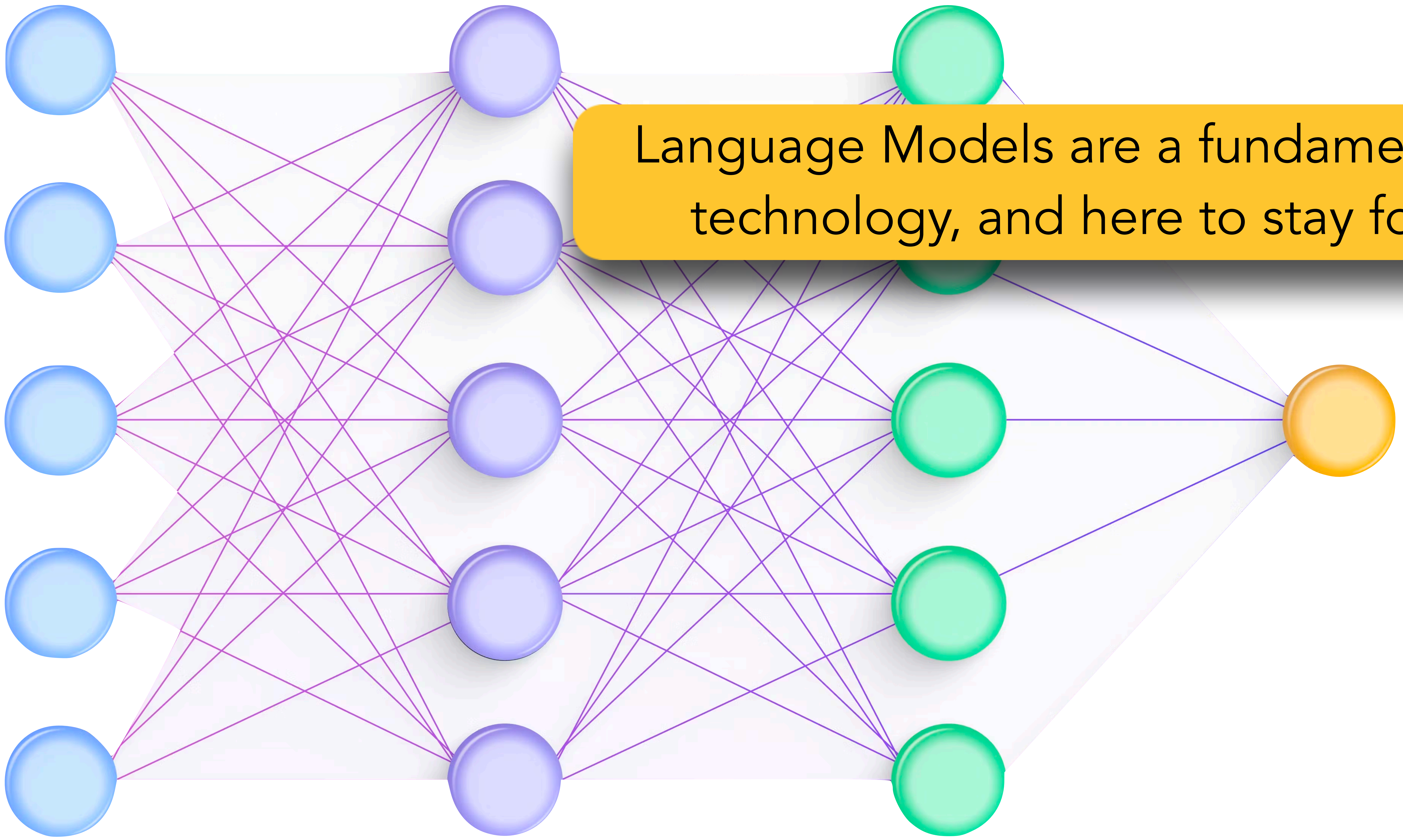
Switch Transformers; Fetus et al., 2021. (https://arxiv.org/abs/2101.03961)

To wrap it up…

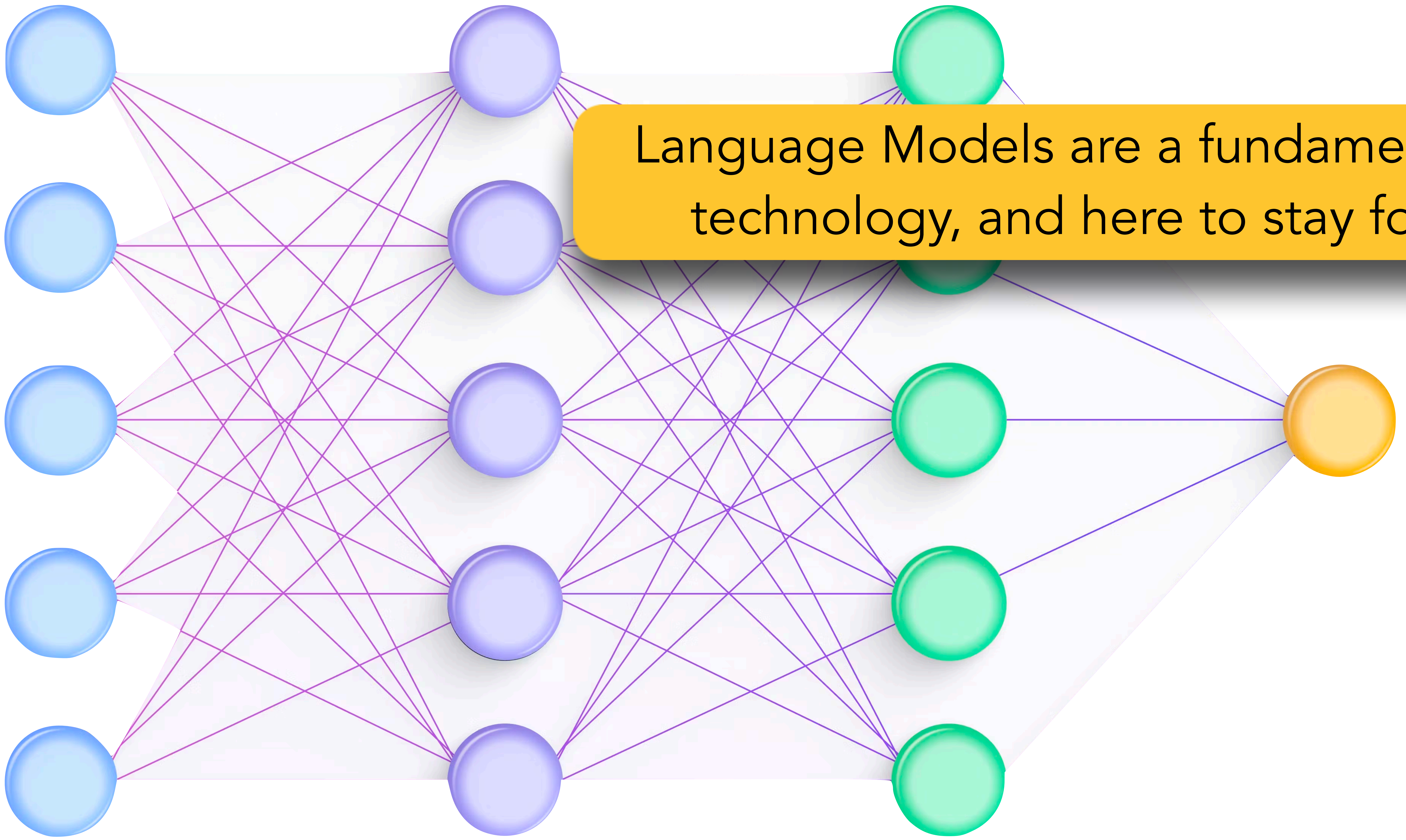
Recap: Learning Objectives

This course is designed to give students an overview of language models in the context of natural language processing. Students will get hands-on experience on developing and evaluating language models trained on (noisy and) real data via class programming assignments. Moreover, students are expected to come away with skills on classical and current NLP practices, as well as communicating their ideas.





Language Models are a fundamental and foundational technology, and here to stay for a long long time!



Language Models are a fundamental and foundational technology, and here to stay for a long long time!

Be creative and ask the important questions as you use this technology

Other Resources

Other Resources

- Fall 2024 classes at USC
  - CSCI 544 by myself - Applied Natural Language Processing: https://swabhs.com/new_teaching/
    - Revamped Syllabus: Focus on language models and more advanced topics
  - CSCI 699 by Robin Jia - Special Topics on Large Language Models

Other Resources

- Fall 2024 classes at USC
  - CSCI 544 by myself - Applied Natural Language Processing: https://swabhs.com/new_teaching/
    - Revamped Syllabus: Focus on language models and more advanced topics
  - CSCI 699 by Robin Jia - Special Topics on Large Language Models
- Other Institutes
  - ETH Zürich - Large Language Models: https://rycolab.io/classes/llm-s23/
  - Stanford - Large Language Models: https://stanford-cs324.github.io/winter2022/

Other Resources

USC Viterbi School of Engineering

- Fall 2024 classes at USC
  - CSCI 544 by myself - Applied Natural Language Processing: https://swabhs.com/new_teaching/
    - Revamped Syllabus: Focus on language models and more advanced topics
  - CSCI 699 by Robin Jia - Special Topics on Large Language Models
- Other Institutes
  - ETH Zürich - Large Language Models: https://rycolab.io/classes/llm-s23/
  - Stanford - Large Language Models: https://stanford-cs324.github.io/winter2022/
- Constantly evolving field
  - Keep up via Twitter and other social media (but be cautious!)
    - e.g. Very accessible LM tutorial: https://www.youtube.com/watch?v=k9DnQPrfJQs&ab_channel=HarvardDataScienceInitiative

Thank You! Go ahead and generate…

Thank You! Go ahead and generate…