# CS 839: Foundation Models
## **Course Overview**

Fred Sala

University of Wisconsin-Madison

**Sept. 7, 2023**

# **Logistics**: Lecture Location

- In-person in **CS 1221**
  - Will have slides / blackboard usage
  - Blackboard for theory; slides for model diagrams etc.

- Planning to record---final decision TBD.

# **Logistics**: Enrollment

- Currently at capacity, approx. 90 students
  - Some folks on waitlist may not make it in
  - Decent chance many of the waitlist folks will
  - **Sorry** ☹ … will be offered again

# **Logistics**: Teaching Team

Instructor: **Fred Sala**

- Location: CS 5385
- Office Hours: Th. 2:30-4:00 pm / by appointment

TA: **Changho Shin**

- Location: CS 3294
- Office Hours: Fri. 2:00-3:00 pm / by appointment

- Note: times possibly **subject to change**

# **Logistics**: Teaching Team

Two more assistants:

**Felix** + **Arthur**

Note: if I’m late replying to anything, they’re the **cause** ☺

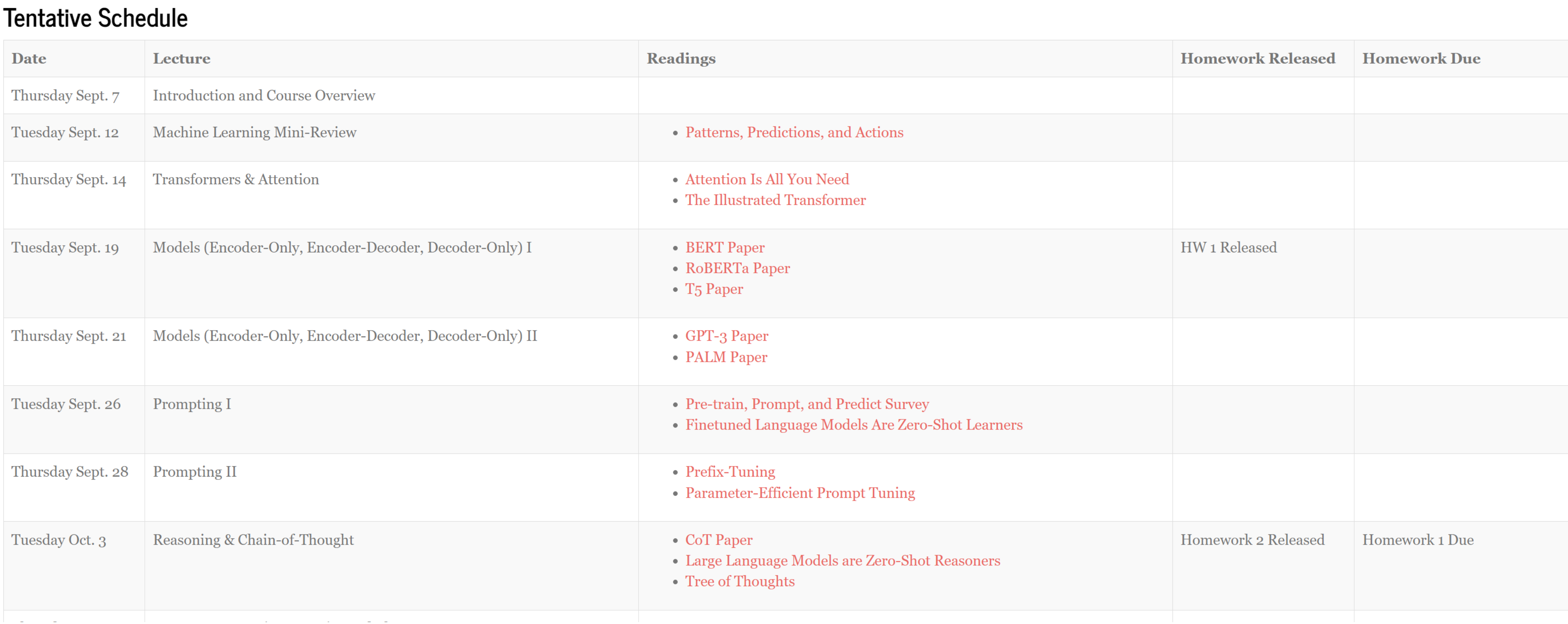
# **Logistics**: Content

Three locations:

- 1. **Course website**: https://pages.cs.wisc.edu/~fredsala/cs839/fall2023/
- 2. **Piazza**. https://piazza.com/class/llfbrbkv5bu15e
  - access code: *introtofm*
  - **Preferred for questions!**
- 3. **Canvas**

Flickr Photo by Jacob Surland - creative commons licensed ( BY-NC )

# Course Content / Schedule



## Tentative Schedule

| Date | Lecture | Readings | Homework Released | Homework Due |
|---|---|---|---|---|
| Thursday Sept. 7 | Introduction and Course Overview | | | |
| Tuesday Sept. 12 | Machine Learning Mini-Review | • Patterns, Predictions, and Actions | | |
| Thursday Sept. 14 | Transformers & Attention | • Attention Is All You Need<br>• The Illustrated Transformer | | |
| Tuesday Sept. 19 | Models (Encoder-Only, Encoder-Decoder, Decoder-Only) I | • BERT Paper<br>• RoBERTa Paper<br>• T5 Paper | HW 1 Released | |
| Thursday Sept. 21 | Models (Encoder-Only, Encoder-Decoder, Decoder-Only) II | • GPT-3 Paper<br>• PALM Paper | | |
| Tuesday Sept. 26 | Prompting I | • Pre-train, Prompt, and Predict Survey<br>• Finetuned Language Models Are Zero-Shot Learners | | |
| Thursday Sept. 28 | Prompting II | • Prefix-Tuning<br>• Parameter-Efficient Prompt Tuning | | |
| Tuesday Oct. 3 | Reasoning & Chain-of-Thought | • CoT Paper<br>• Large Language Models are Zero-Shot Reasoners<br>• Tree of Thoughts | Homework 2 Released | Homework 1 Due |

# **Logistics**: Lecture Formats

Two types of class sessions:

- **Type 1: Lectures**
  - Mostly slides, some whiteboard
  - Will take some breaks, 1-2 during the lecture
  - Can ask questions---during lecture and breaks

- **Type 2: Paper Presentations**
  - More info on later slides.

- Start with Type 1, conclude semester with Type 2

# **Logistics**: Assignments & Grades

**Homeworks**:

- 3 or so, worth 30% total
- Posted after class; due when class starts on due date. About 2-3 weeks given for each one

**Class Presentation**:

- Total of 30%. Present a paper
- Split up into groups of 3-6 students. Proposal midway, check-ins.

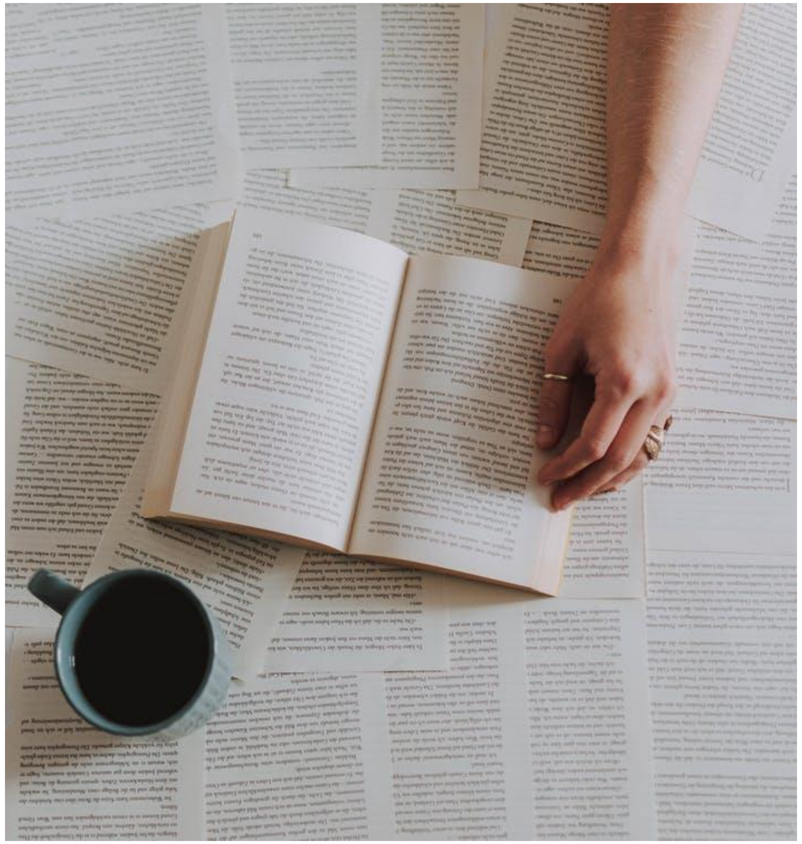
**Final Project**:

- 40% total, groups of 3-6; proposal midway. **More info soon!**

# **Class Setup**: Reading

No textbooks

- I will post useful notes, primers, papers
- Expect **new papers** (submitted during the timeframe of the class)
- For presentations: we will have a list of papers to pick from, but new/unlisted papers are options as well



# **Class Setup**: Background

More on this at the end of class, but

- **Basic ML** (at the level of 760 or so)
  - Short review next lecture
- **Technical components:**
  - Linear Algebra
  - Calculus
  - Probability

Note: this class is partially **conceptual** and partially **technical**

# **Class Setup**: Goals

Two goals:

- Become acquainted with **how to use** large pretrained/language/foundation models
- Understanding the technical underpinnings of these models and ***why*** they work

**Note**: if you are only interested in a very broad overview of ML, then CS 540 or 760 might be a better choice.

# **Class Setup**: Goals II

Mini-goals:

- **Understanding** research
- **Big picture/**ML ecosystem
- **Intuition** around modern ML paradigms

# Break & Questions

# What We'll Cover

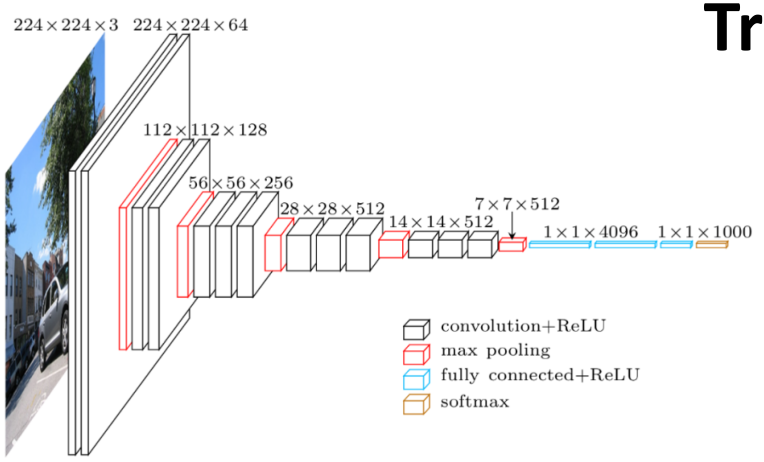
- The past: **supervised learning**
  - **Dataset:** $(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(n)}, y^{(n)})$

safe

poisonous

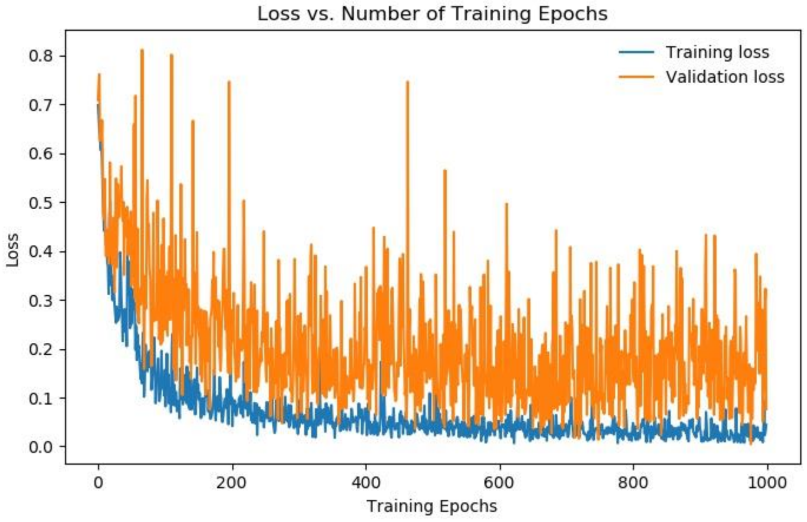
safe

  - **Model:**

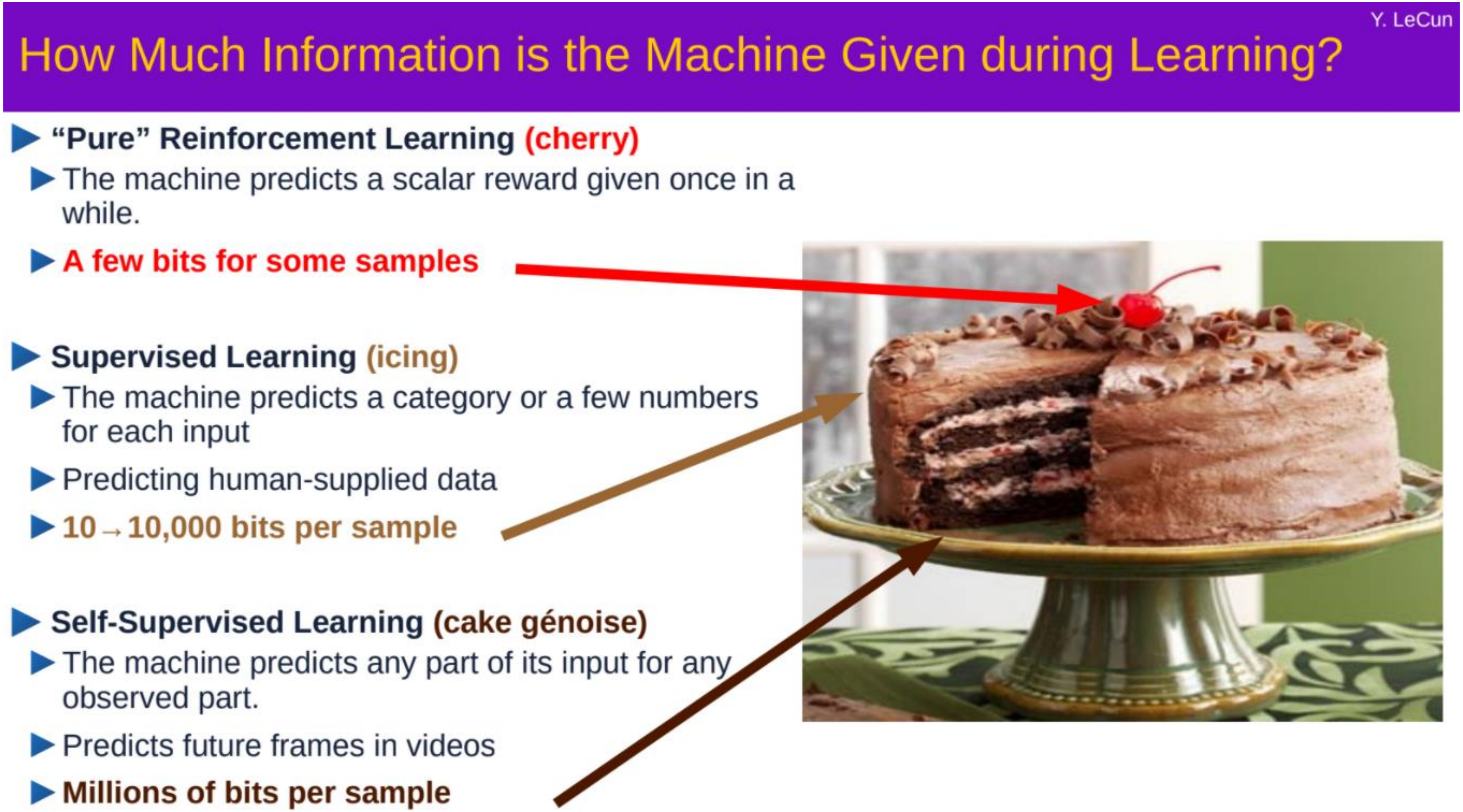


Simonyan and Zisserman

**Train:**



# New Paradigms: Pretraining



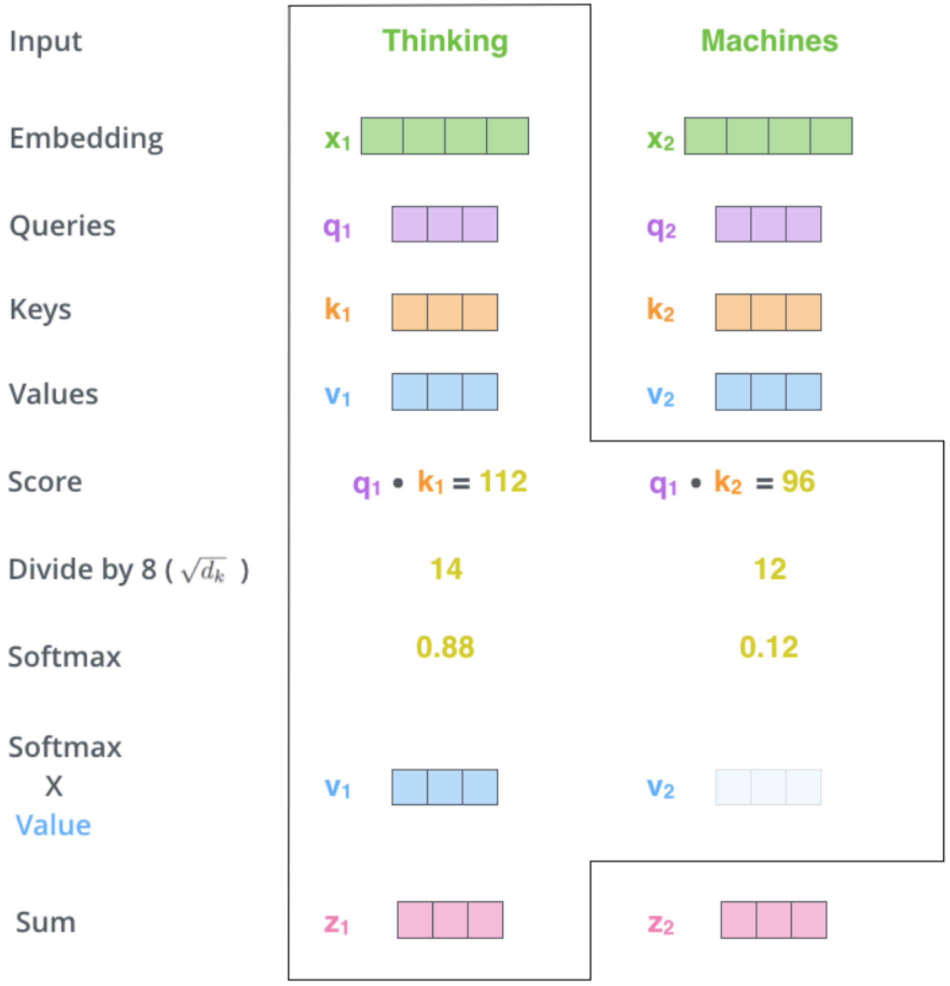
Y. LeCun

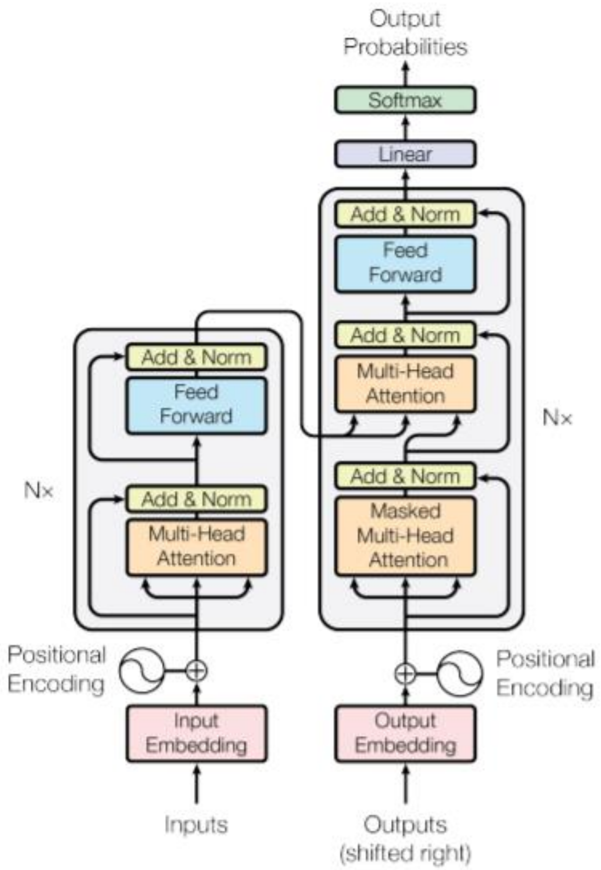
## How Much Information is the Machine Given during Learning?

- **“Pure” Reinforcement Learning (cherry)**
  - The machine predicts a scalar reward given once in a while.
  - **A few bits for some samples**
- **Supervised Learning (icing)**
  - The machine predicts a category or a few numbers for each input
  - Predicting human-supplied data
  - **10→10,000 bits per sample**
- **Self-Supervised Learning (cake génoise)**
  - The machine predicts any part of its input for any observed part.
  - Predicts future frames in videos
  - **Millions of bits per sample**

# New Architectures: Transformers

- Lots of architectures (RNNs, LSTMs, CNNs) → Transformers





Vaswani et al. '17

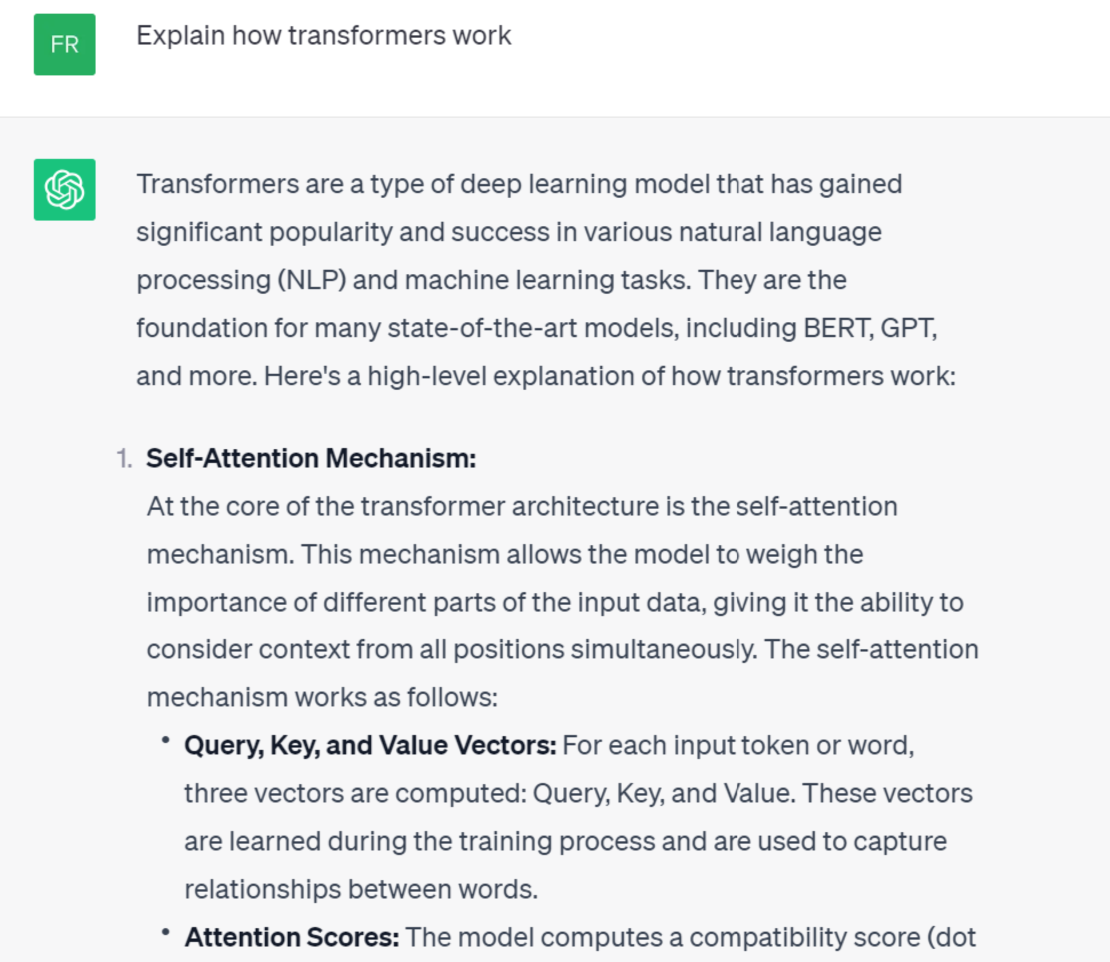
# New Models

A bunch you've seen already,

GPT−3 BERT T5 BLOOM stability.ai

And more all the time:

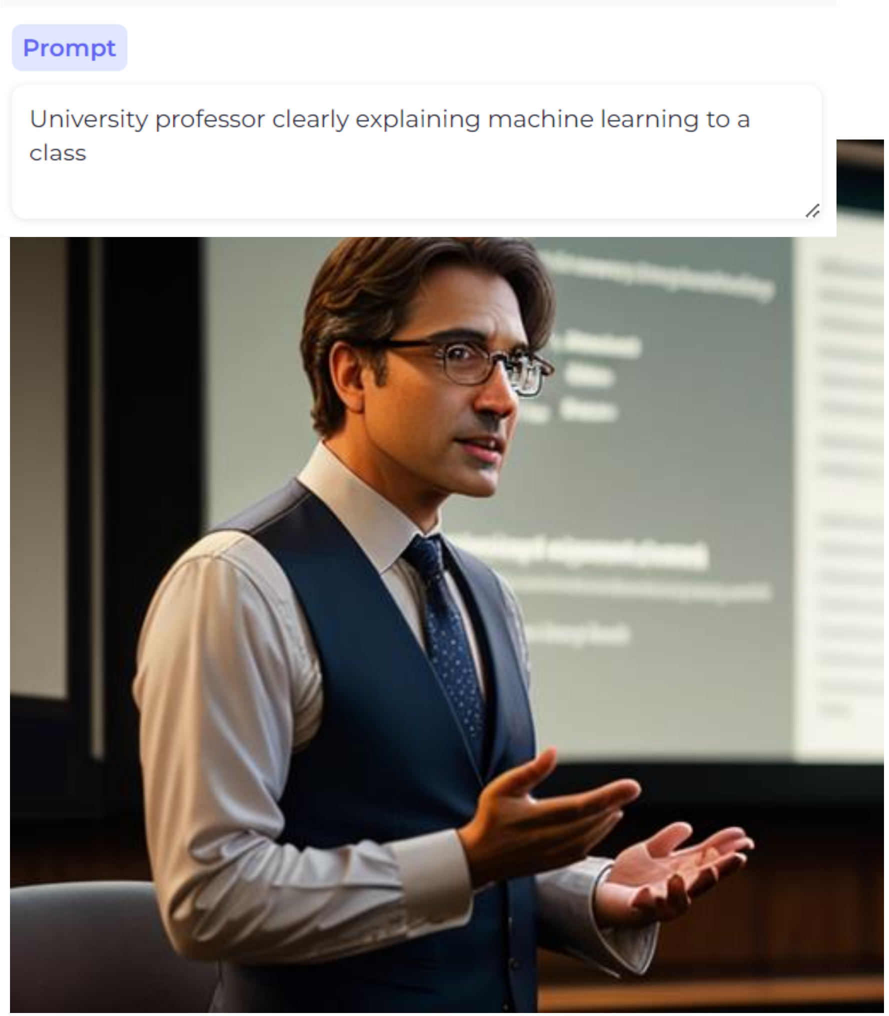
# New Interfaces: Prompting



Explain how transformers work

Transformers are a type of deep learning model that has gained significant popularity and success in various natural language processing (NLP) and machine learning tasks. They are the foundation for many state-of-the-art models, including BERT, GPT, and more. Here's a high-level explanation of how transformers work:

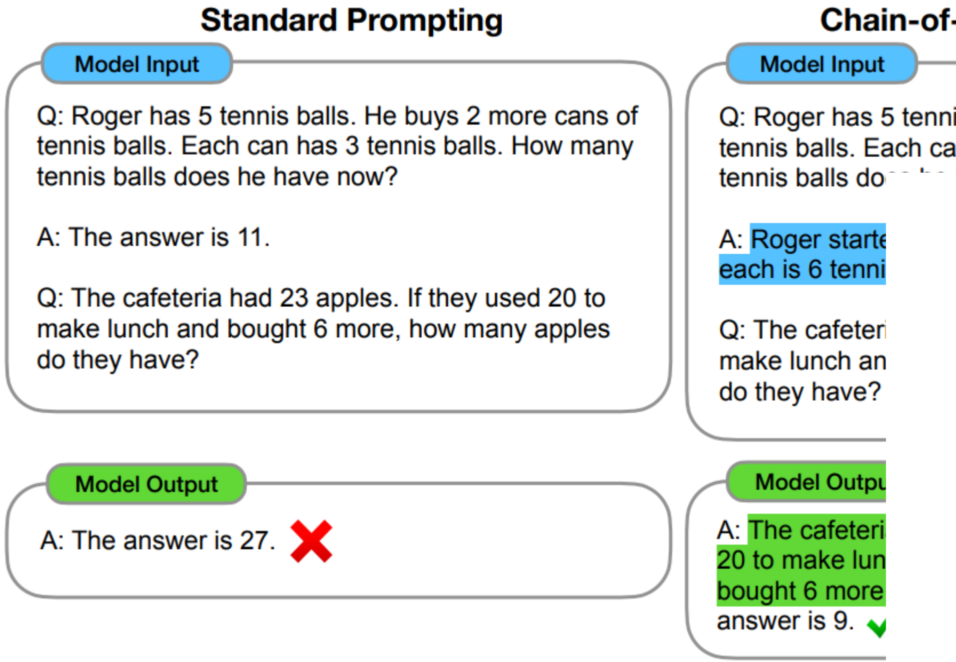
1. **Self-Attention Mechanism:**
   At the core of the transformer architecture is the self-attention mechanism. This mechanism allows the model to weigh the importance of different parts of the input data, giving it the ability to consider context from all positions simultaneously. The self-attention mechanism works as follows:
   - **Query, Key, and Value Vectors:** For each input token or word, three vectors are computed: Query, Key, and Value. These vectors are learned during the training process and are used to capture relationships between words.
   - **Attention Scores:** The model computes a compatibility score (dot



University professor clearly explaining machine learning to a class

# Reasoning

- Chain-of-thought and friends:



**Standard Prompting**

**Model Input**

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

**Model Output**

A: The answer is 27. ✗

**Chain-of-Thought Prompting**

**Model Input**

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls do

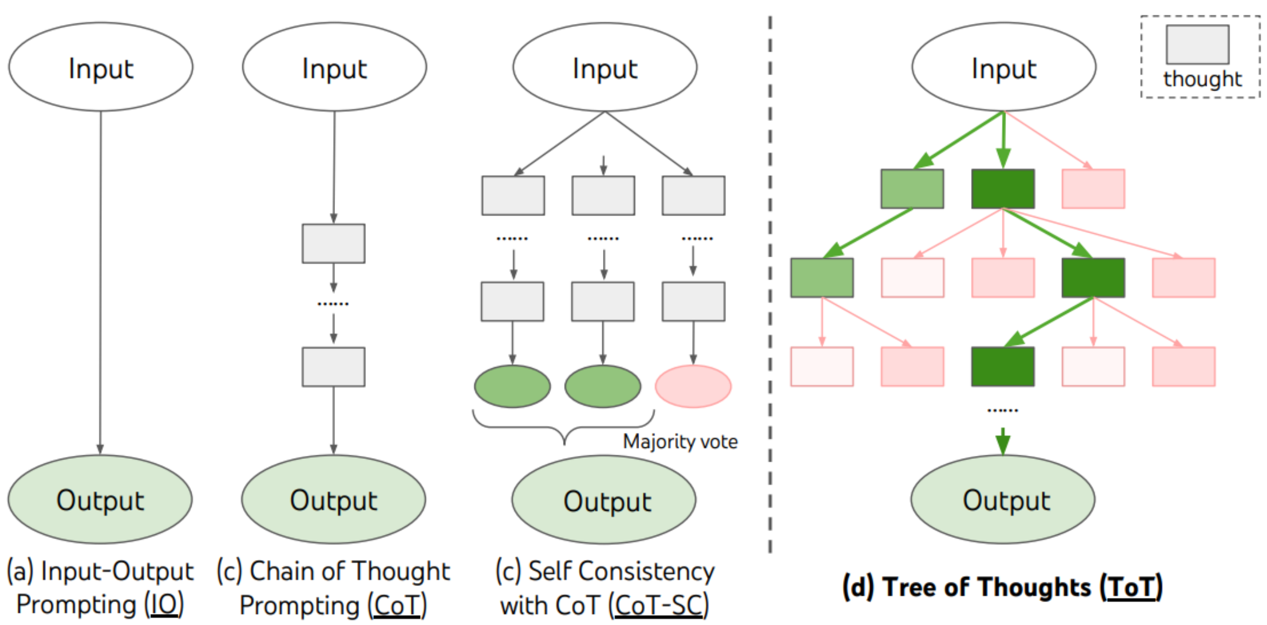
A: Roger starte
each is 6 tenni

Q: The cafeteri
make lunch an
do they have?

**Model Outpu**

A: The cafeteri
20 to make lun
bought 6 more
answer is 9. ✓

Wei et al



(a) Input-Output Prompting (IO)

(c) Chain of Thought Prompting (CoT)

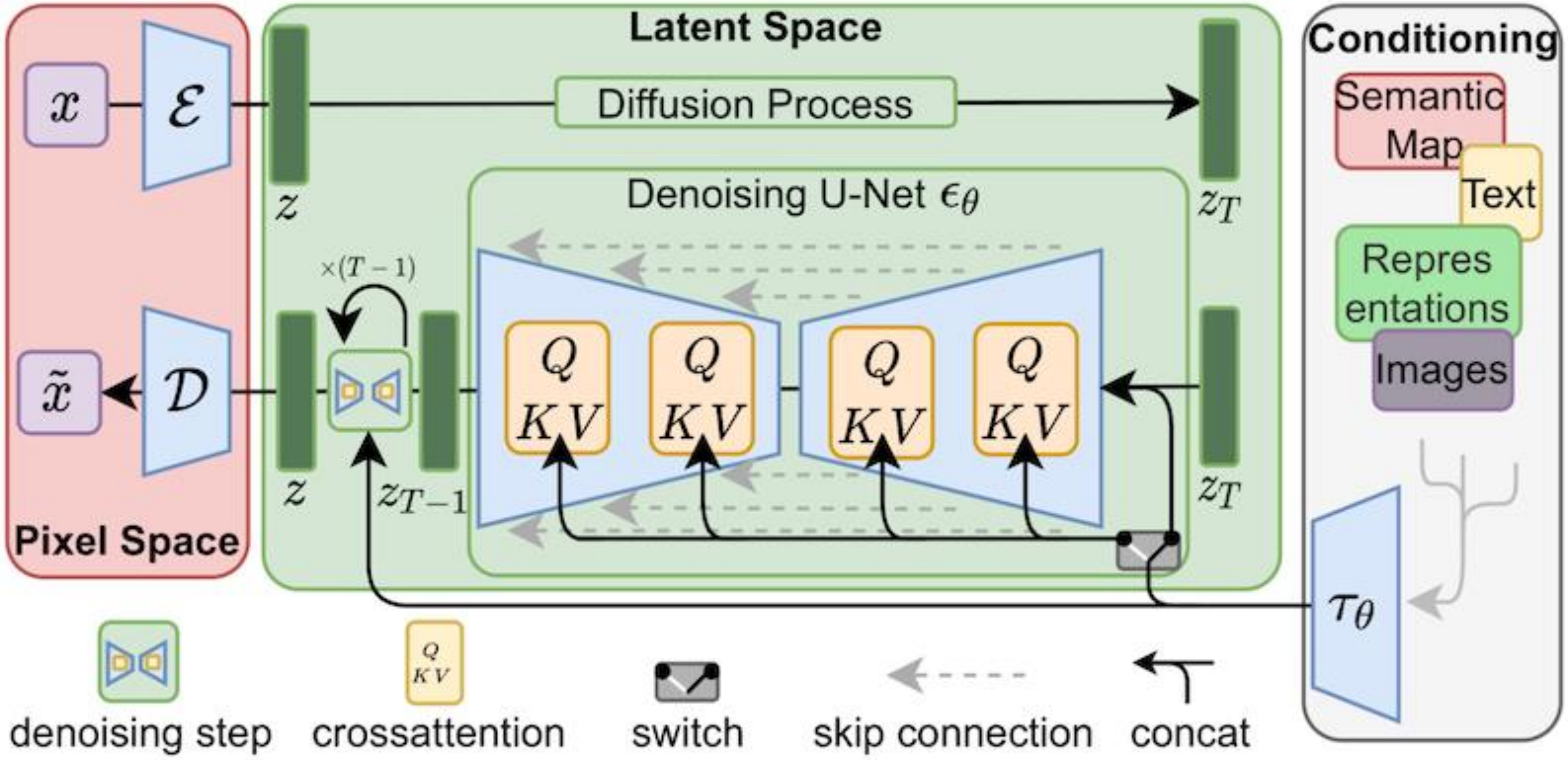
(c) Self Consistency with CoT (CoT-SC)

**(d) Tree of Thoughts (ToT)**

Yao et al

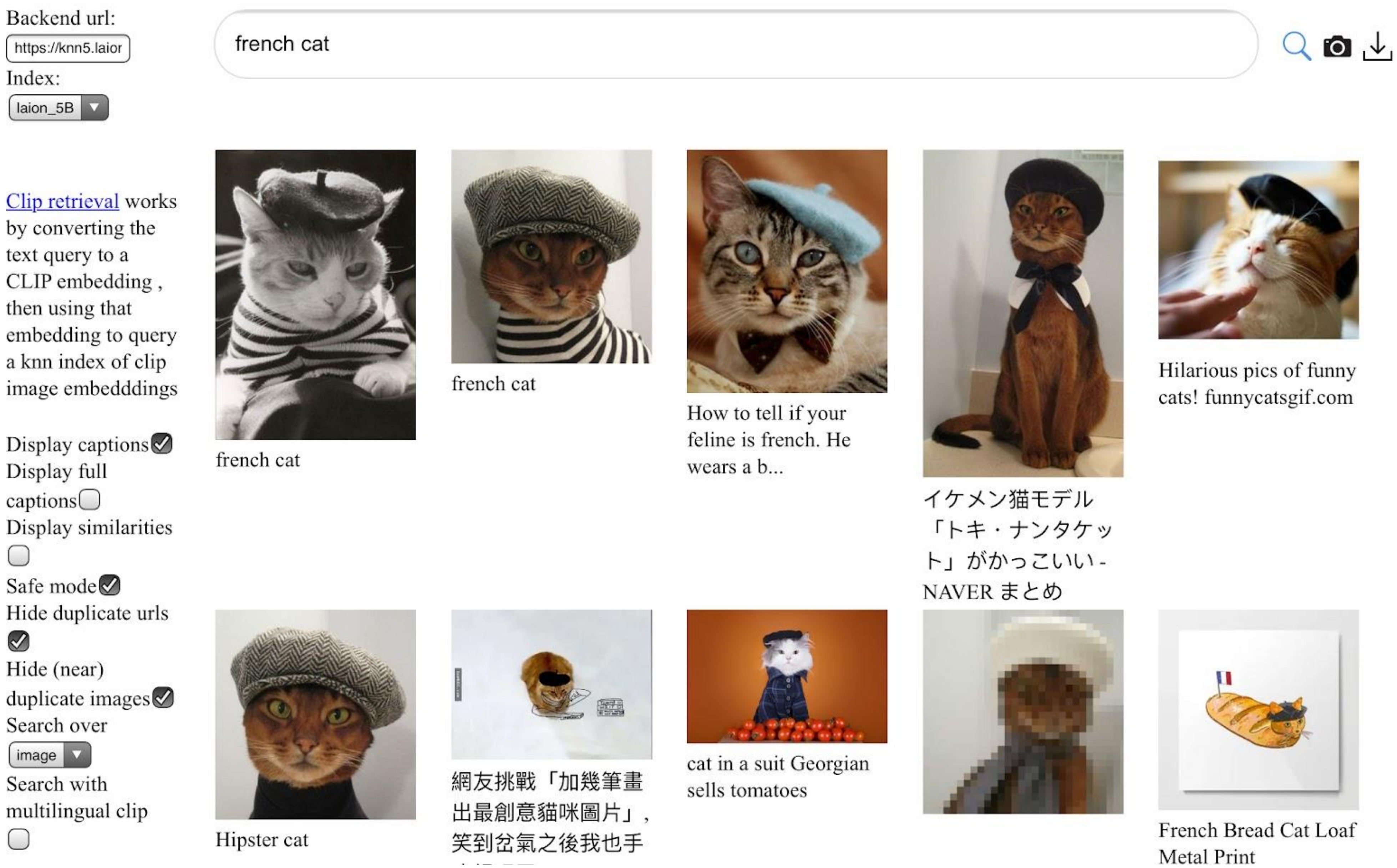
# Adapting & Improving Models

- Prompt Engineering
- Fine-tuning
- Adaptation



Cuenca and Paul

# Training & Data



Backend url: https://knn5.laior

Index: laion_5B

french cat

[Clip retrieval](#) works by converting the text query to a CLIP embedding , then using that embedding to query a knn index of clip image embedddings

Display captions ☑

Display full captions ☐

Display similarities ☐

Safe mode ☑

Hide duplicate urls ☑

Hide (near) duplicate images ☑

Search over image

Search with multilingual clip ☐

french cat

french cat

How to tell if your feline is french. He wears a b...

イケメン猫モデル「トキ・ナンタケット」がかっこいい - NAVER まとめ

Hilarious pics of funny cats! funnycatsgif.com

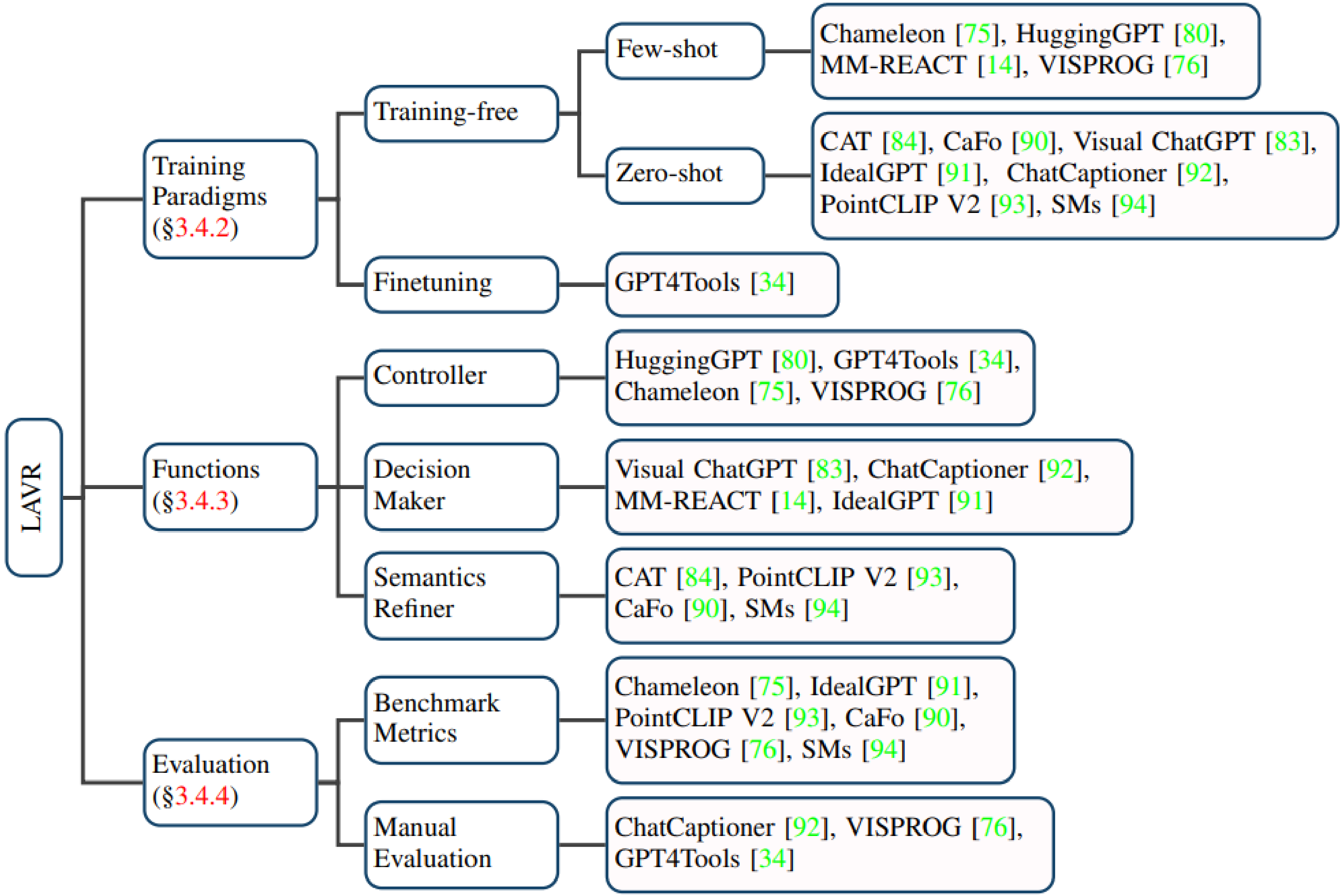
Hipster cat

網友挑戰「加幾筆畫出最創意貓咪圖片」，笑到岔氣之後我也手

cat in a suit Georgian sells tomatoes

French Bread Cat Loaf Metal Print

# Multimodal Models

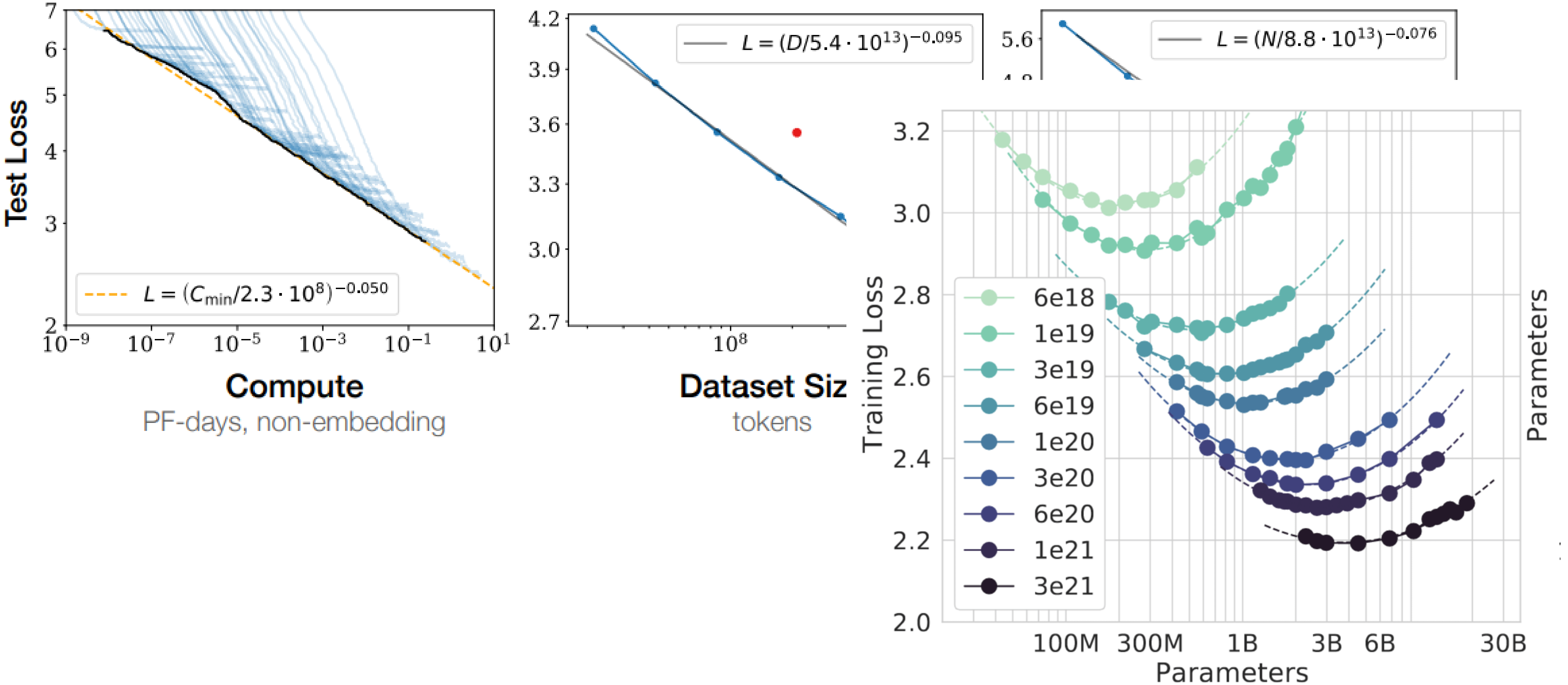


- LAVR
  - Training Paradigms (§3.4.2)
    - Training-free
      - Few-shot
        - Chameleon [75], HuggingGPT [80], MM-REACT [14], VISPROG [76]
      - Zero-shot
        - CAT [84], CaFo [90], Visual ChatGPT [83], IdealGPT [91], ChatCaptioner [92], PointCLIP V2 [93], SMs [94]
    - Finetuning
      - GPT4Tools [34]
  - Functions (§3.4.3)
    - Controller
      - HuggingGPT [80], GPT4Tools [34], Chameleon [75], VISPROG [76]
    - Decision Maker
      - Visual ChatGPT [83], ChatCaptioner [92], MM-REACT [14], IdealGPT [91]
    - Semantics Refiner
      - CAT [84], PointCLIP V2 [93], CaFo [90], SMs [94]
  - Evaluation (§3.4.4)
    - Benchmark Metrics
      - Chameleon [75], IdealGPT [91], PointCLIP V2 [93], CaFo [90], VISPROG [76], SMs [94]
    - Manual Evaluation
      - ChatCaptioner [92], VISPROG [76], GPT4Tools [34]

Yin et al

# Scaling

Scaling laws:



# Security, Privacy, Bias

Some of the issues we’ll encounter…

RESEARCH 03.24.2023

CYBERSECURITY

## THE DARK SIDE OF LARGE LANGUAGE MODELS

### Part 2: “Who’s a good chatbot?”

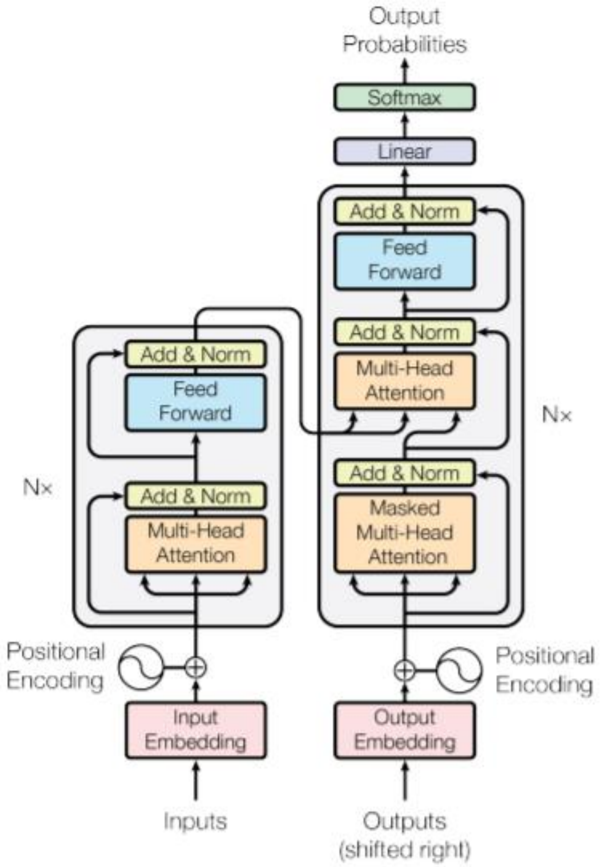
By: Eoin Wickens, Marta Janus

# Next two weeks

1. **Review of ML**
    - Very short!

2. **Architectures: Transformers**
    - Intro to attention.

3. **Language Models**
    - Encoder-decoder, Encoder-only, Decoder-only, etc



Vaswani et al. ‘17