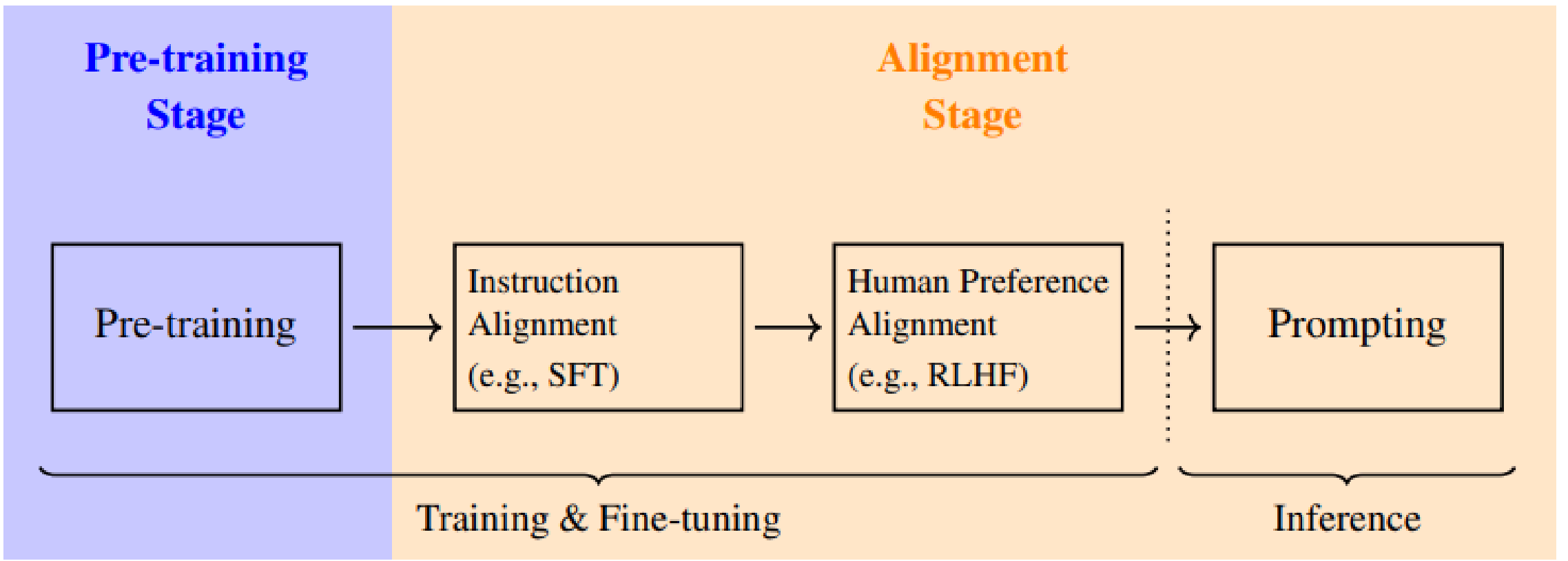
LLM Foundations

1/16

Overview

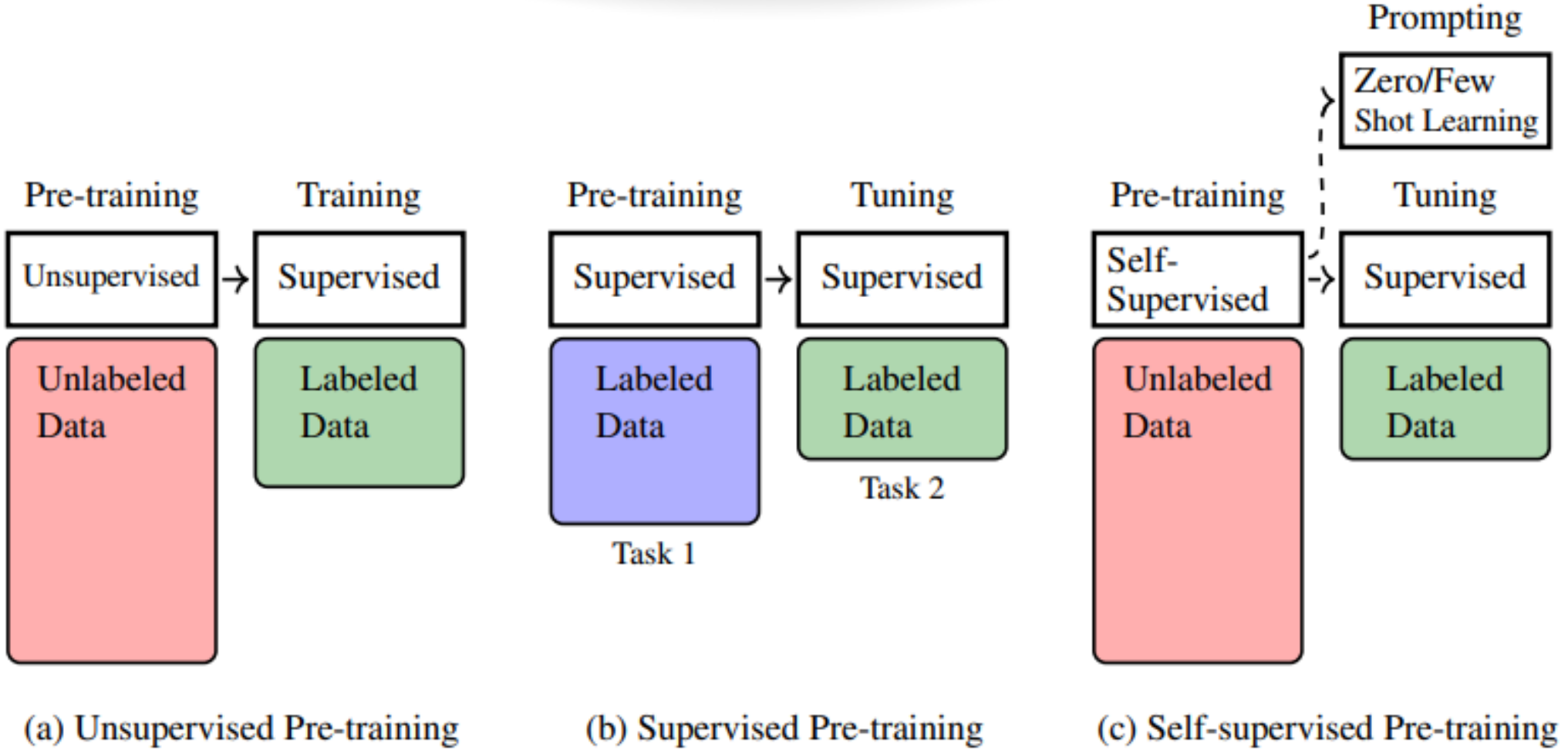
Curated by:
Dr. Maryam Miradi



LLM Foundations

2/16

Pre-training types



(a) Unsupervised Pre-training (b) Supervised Pre-training (c) Self-supervised Pre-training

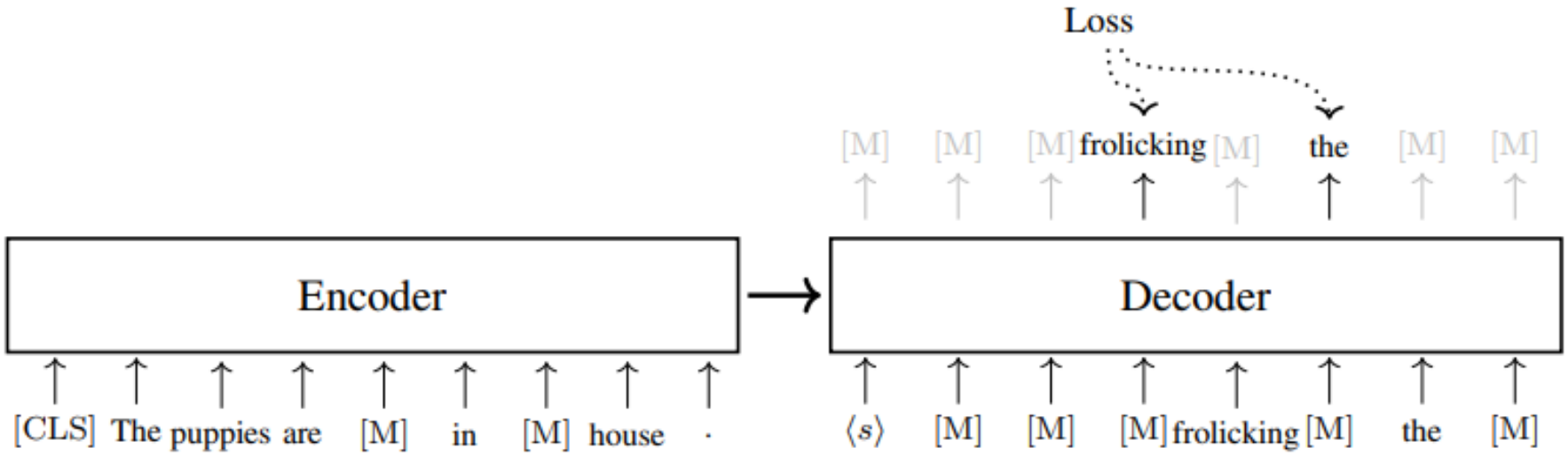
MARYAM MIRADI

Curated by:
Dr. Maryam Miradi

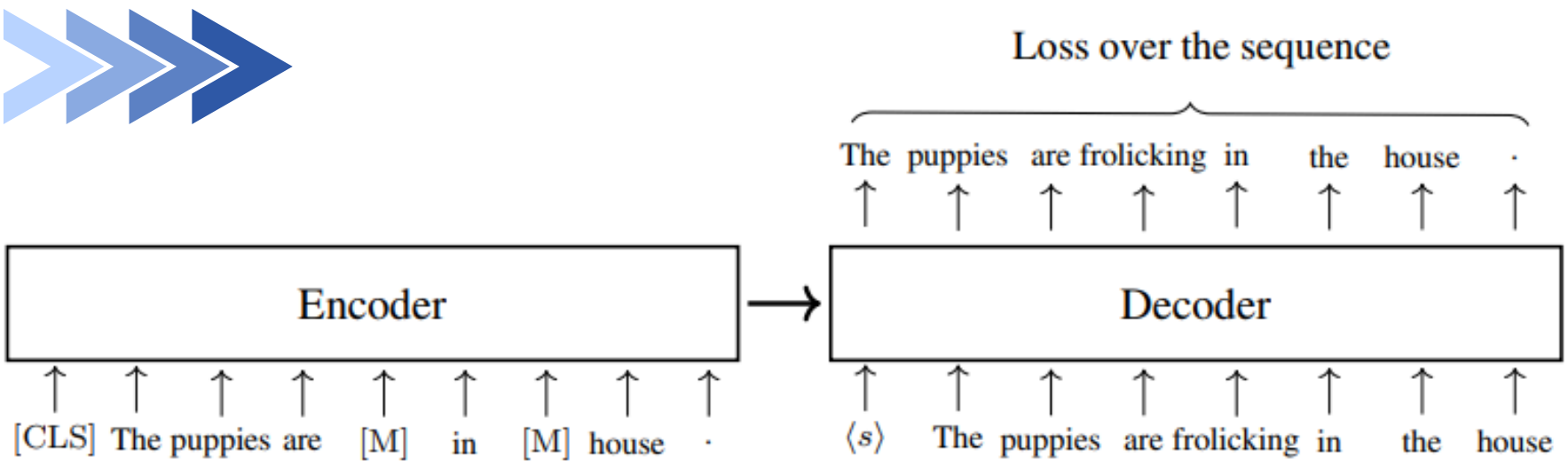
LLM Foundations

3/16

Encoders &Decoders



(a) Training an encoder-decoder model with BERT-style masked language modeling



(b) Training an encoder-decoder model with denoising autoencoding

MARYAM MIRADI

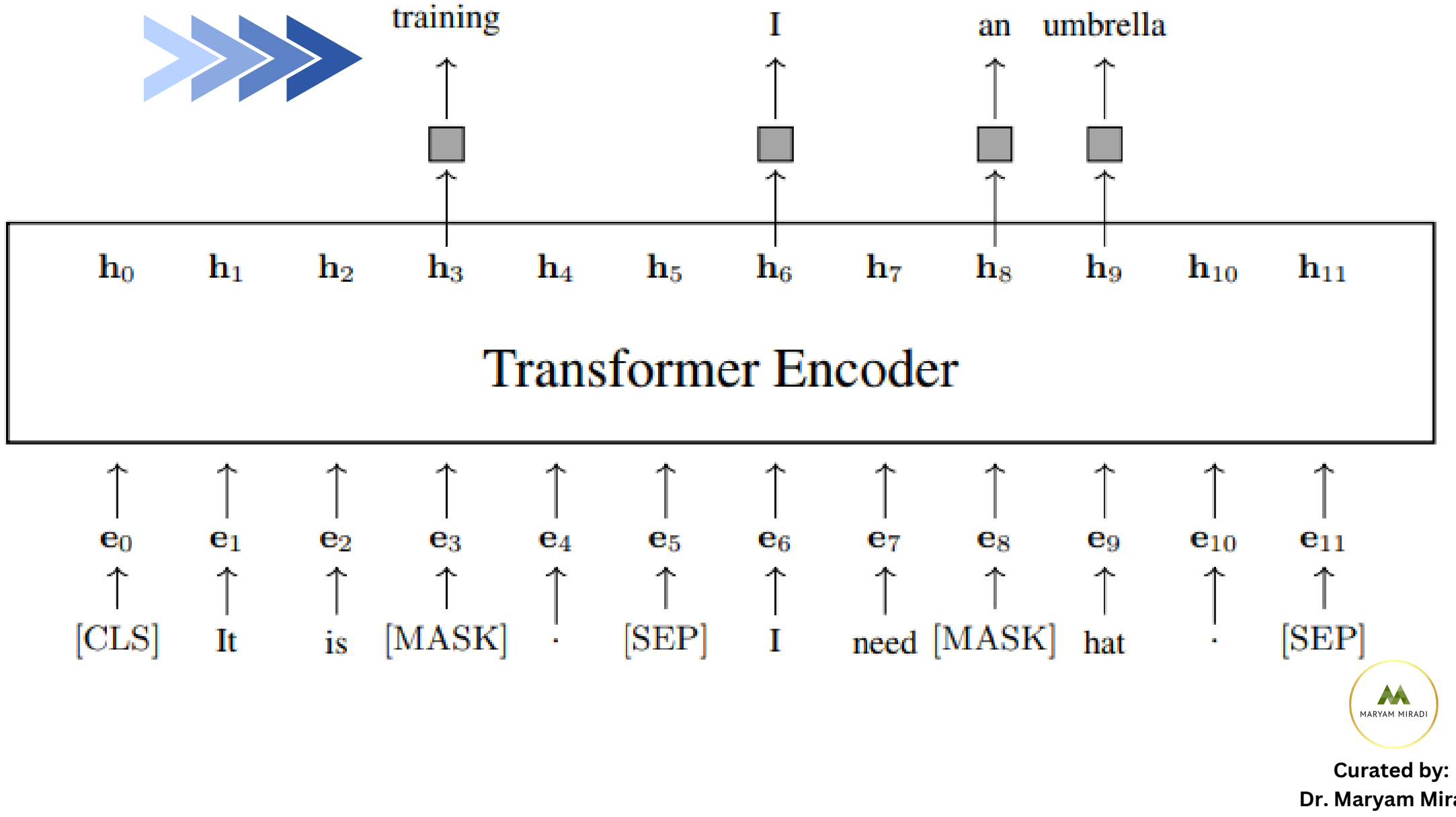
Curated by:
Dr. Maryam Miradi

4/16

LLM Foundations

Transformers

Pre-training



Curated by:
Dr. Maryam Miradi

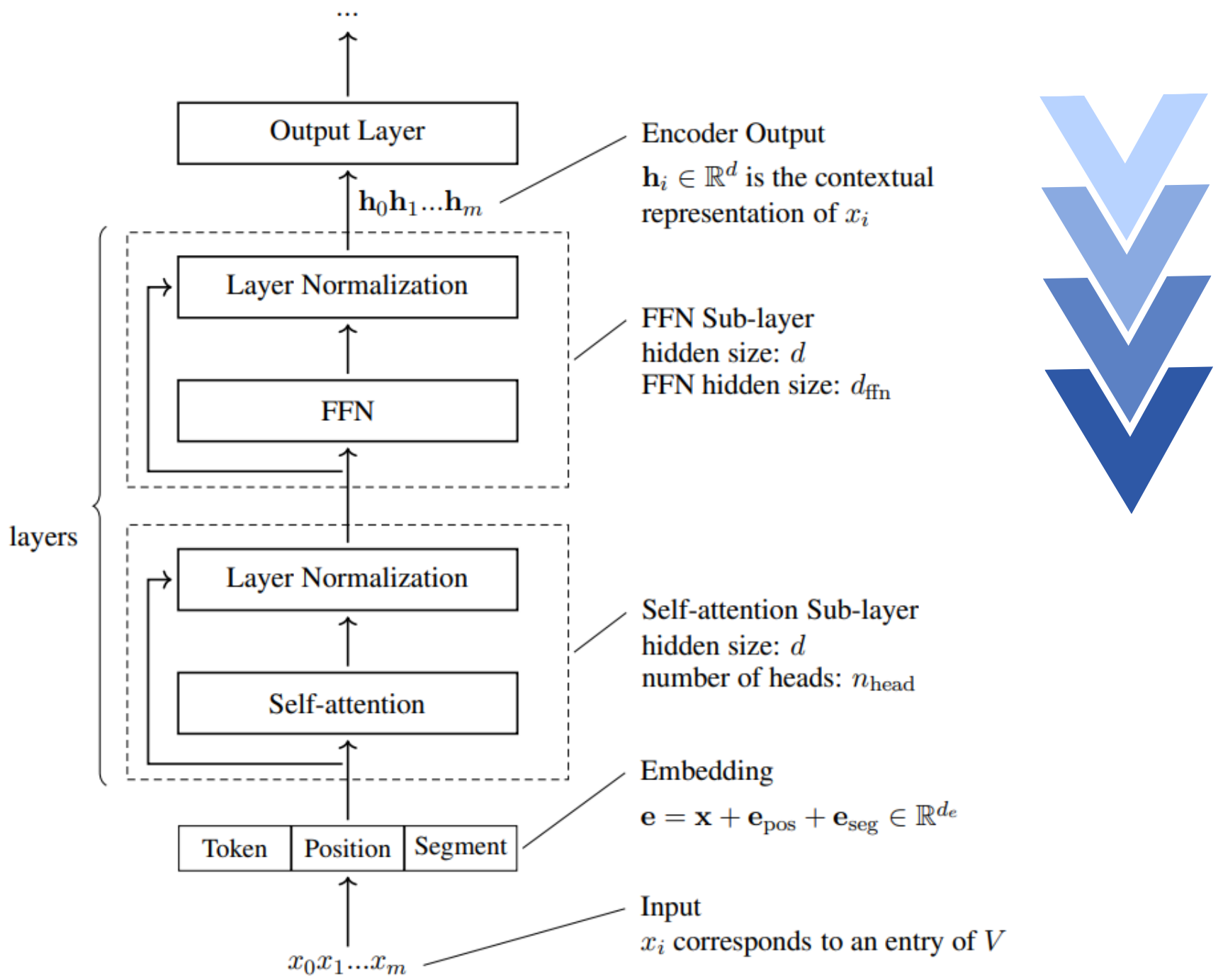
LLM Foundations

5/16

BERT

MARYAM MIRADI

Curated by:
Dr. Maryam Miradi



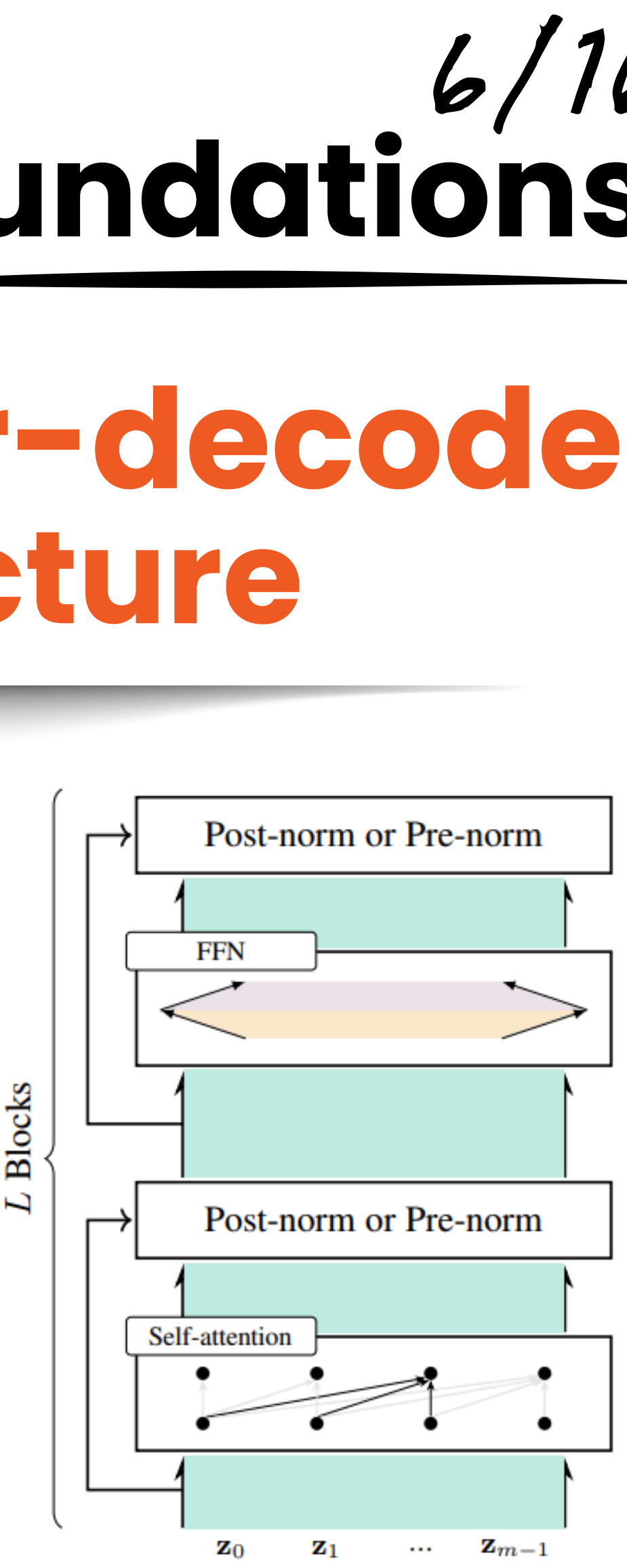
LLM Foundations

6/16

Transformer-decoder architecture

MARYAM MIRADI

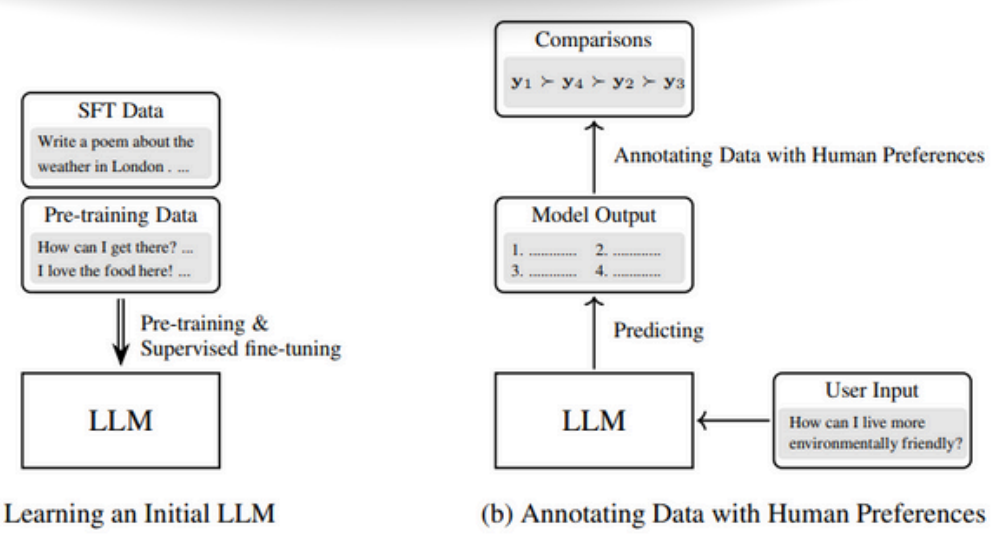
Curated by:
Dr. Maryam Miradi



LLM Foundations

7/16

RLHF



MARYAM MIRADI

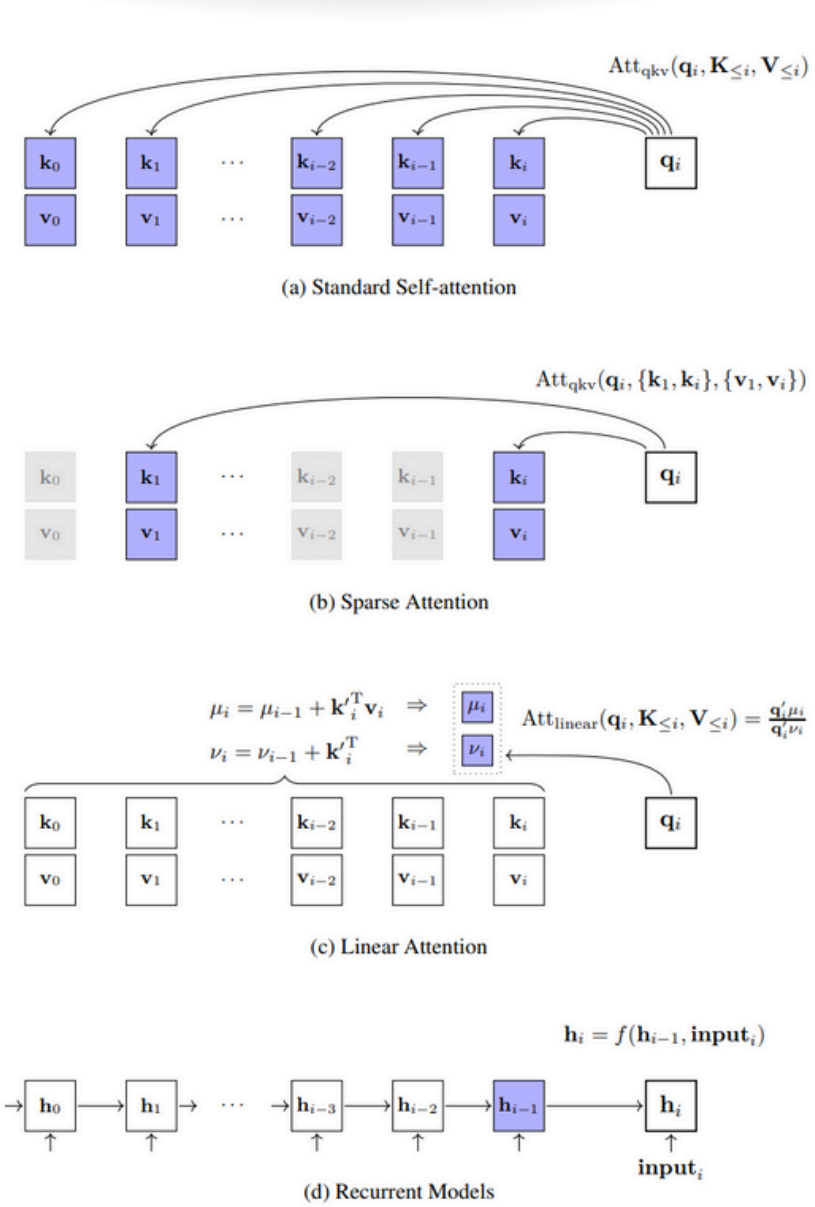
Curated by:
Dr. Maryam Miradi

8/16

LLM Foundations

Attention

Types



(a) Standard Self-attention

(b) Sparse Attention

(c) Linear Attention

(d) Recurrent Models

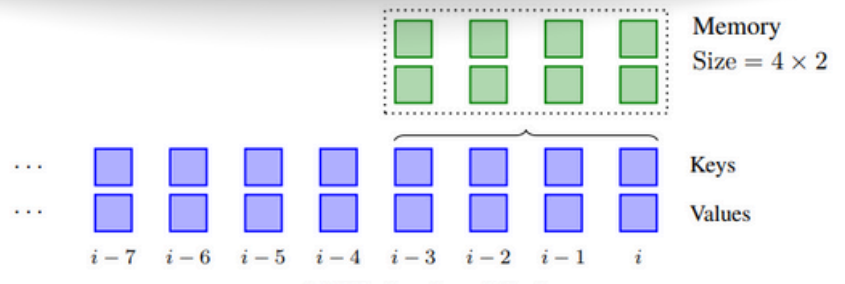
MARYAM MIRADI

Curated by:
Dr. Maryam Miradi

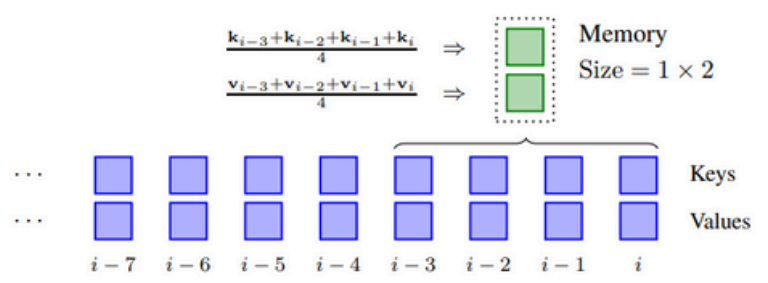
LLM Foundations

9/16

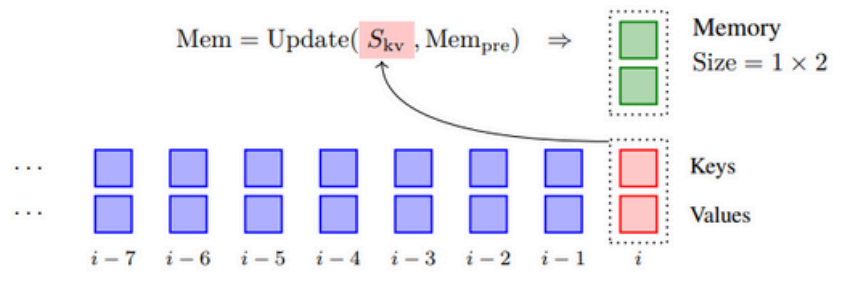
Memory



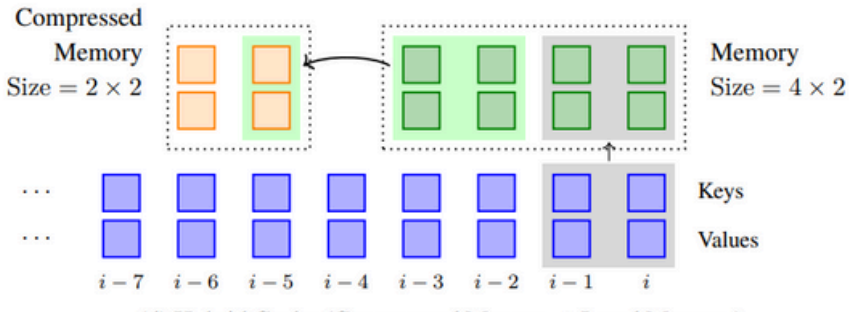
(a) Window-based Cache



(b) Moving Average-based Cache



(c) Recurrent Network as Cache



(d) Hybrid Cache (Compressed Memory + Local Memory)

MARYAM MIRADI

Curated by:
Dr. Maryam Miradi

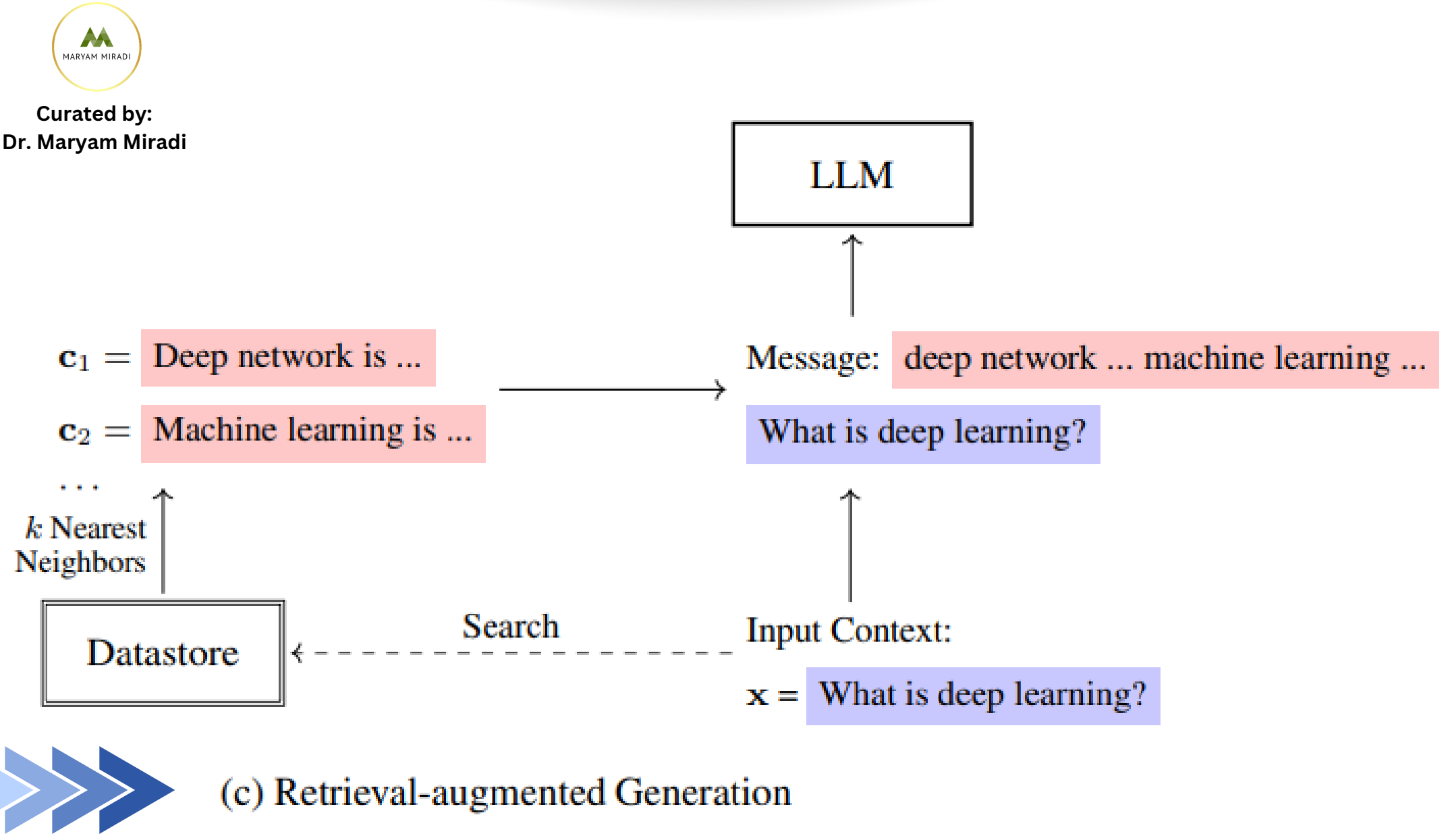
LLM Foundations

10/16

RAG

MARYAM MIRADI

Curated by:
Dr. Maryam Miradi

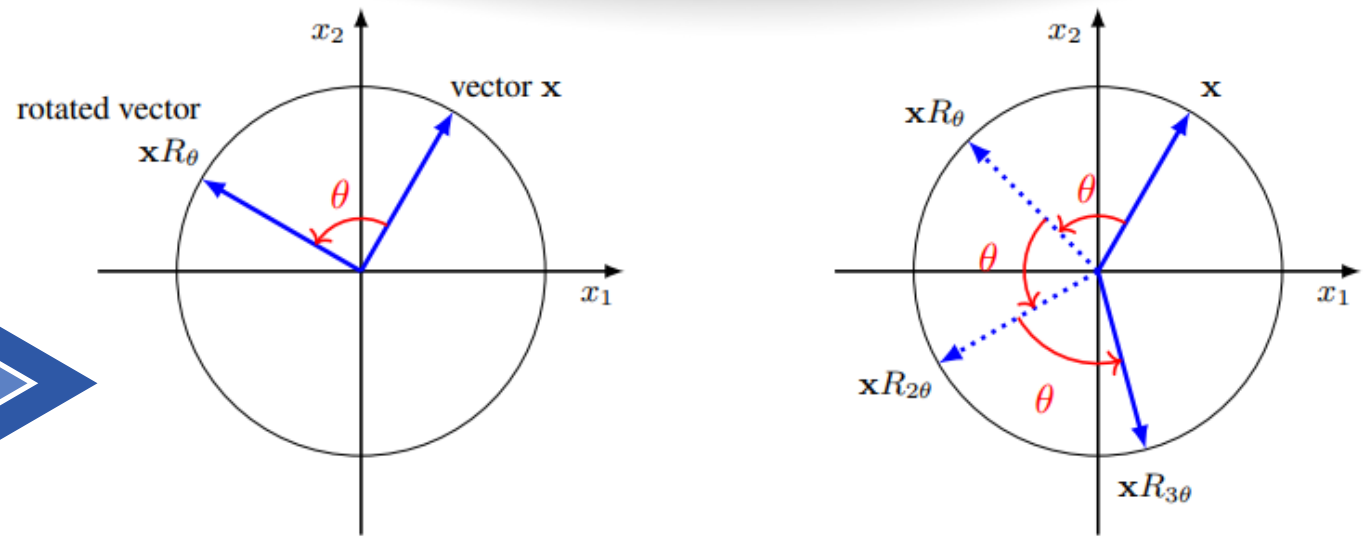


(c) Retrieval-augmented Generation

LLM Foundations

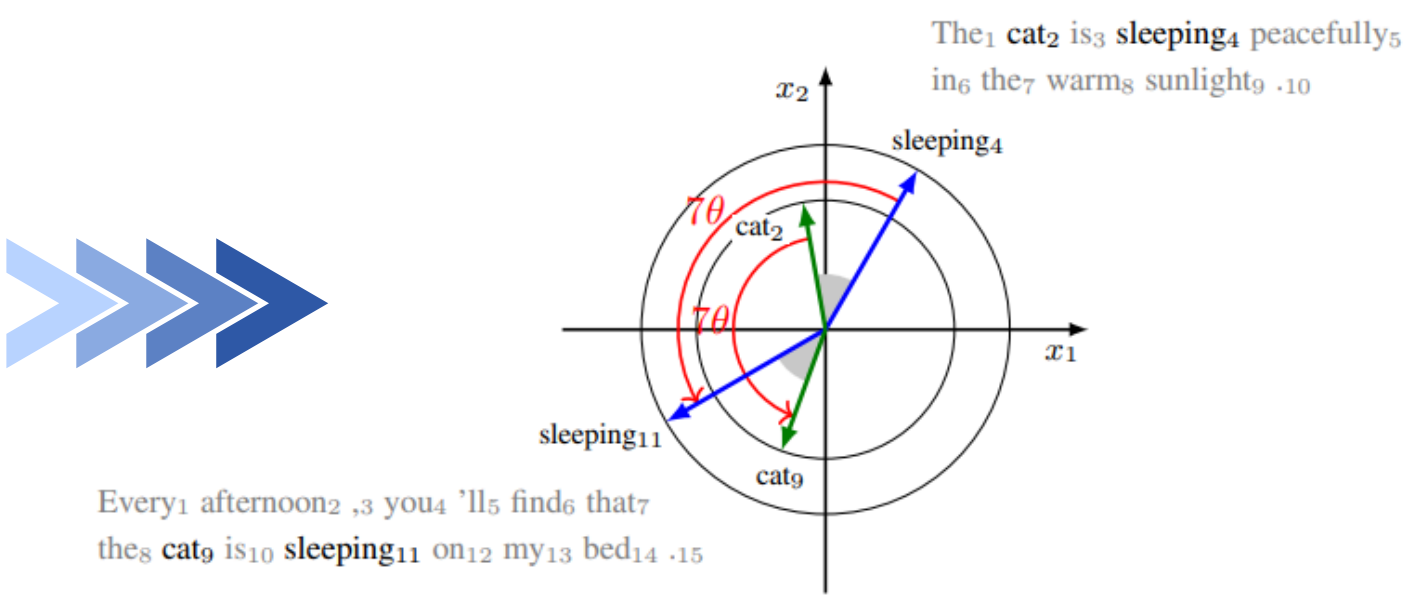
11/16

Embedding



(a) Single-step Rotation

(b) Multi-step Rotation



(c) Angles between embeddings of two tokens at different positions

MARYAM MIRADI

Curated by:
Dr. Maryam Miradi

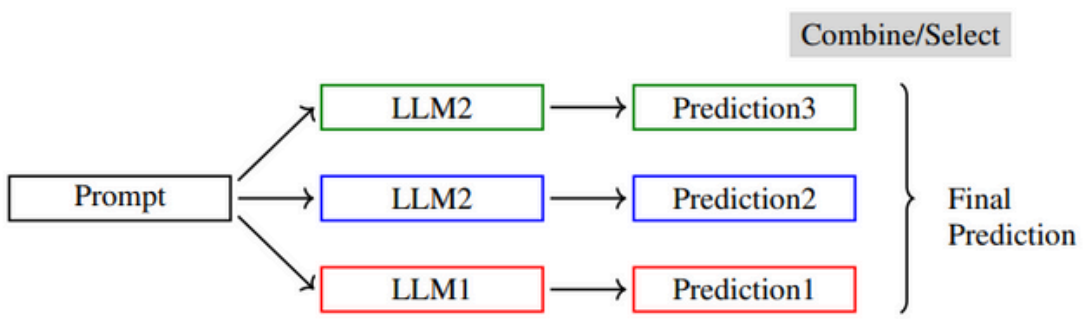
LLM Foundations

12/16

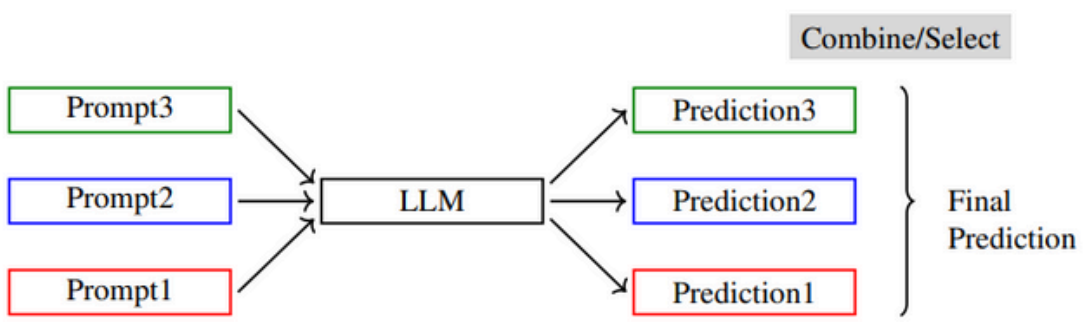
Ensembling

MARYAM MIRADI

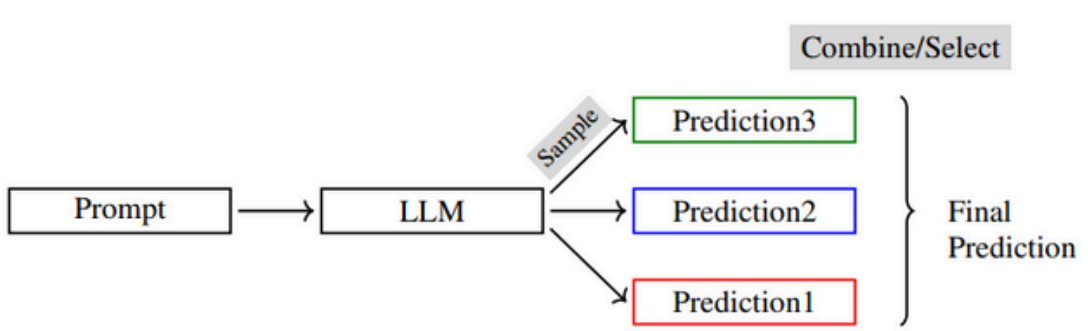
Curated by:
Dr. Maryam Miradi



(a) Model Ensembling



(b) Prompt Ensembling



(c) Output Ensembling

LLM Foundations

13/16

Fine-Tunning

$\text{Loss} = 0$ $\text{Loss} \neq 0$ $\text{Loss} = 0$ $\text{Loss} \neq 0$

$\Pr_\theta(\mathbf{x}^1)$ $\Pr_\theta(\mathbf{y}^1|\mathbf{x}^1)$ $\Pr_\theta(\mathbf{x}^2|\mathbf{x}^1,\mathbf{y}^1)$ $\Pr_\theta(\mathbf{y}^2|\mathbf{x}^1,\mathbf{y}^1,\mathbf{x}^2)$

$\mathbf{x}^1$ $\mathbf{y}^1$ $\mathbf{x}^2$ $\mathbf{y}^2$ ...

User: I've been feeling very tired lately.

Chatbot: I'm sorry to hear that. Besides feeling tired, have you noticed any other symptoms?

User: Yes, I'm also experiencing headaches frequently.

Chatbot: How long have these symptoms been going on?

⋮

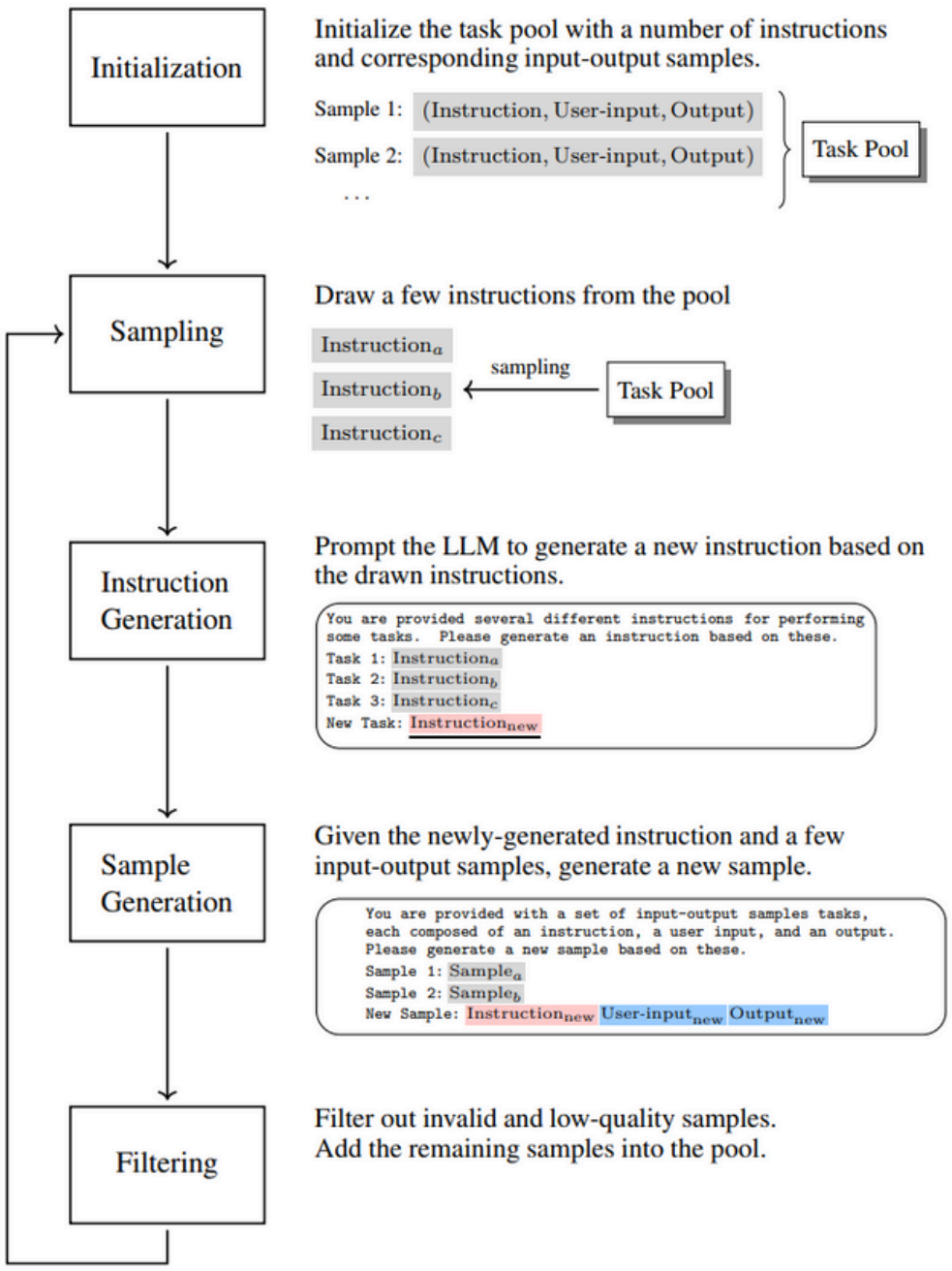
MARYAM MIRADI

Curated by:
Dr. Maryam Miradi

LLM Foundations

14/16

Self-Instruct



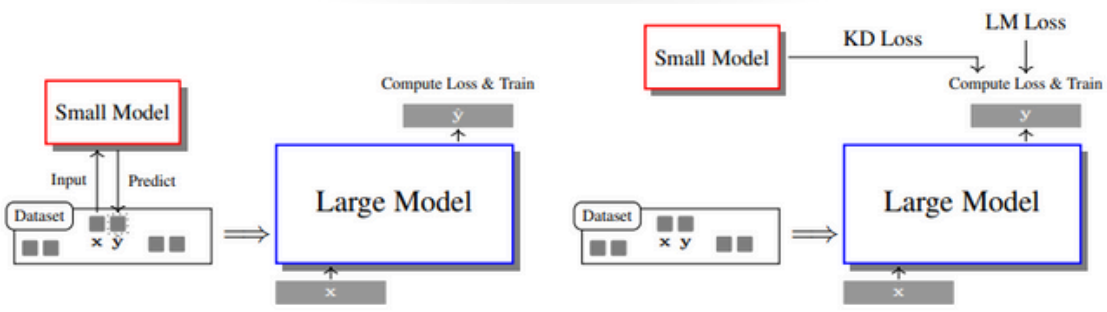
MARYAM MIRADI

Curated by:
Dr. Maryam Miradi

15/16

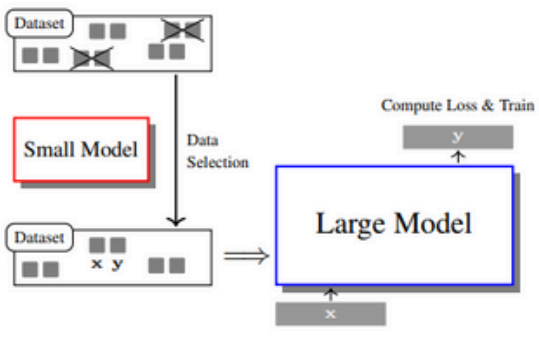
LLM Foundations

Small-to-Large

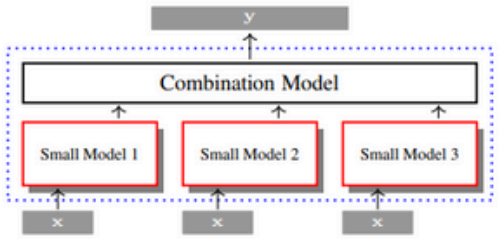


(a) Fine-tuning on data generated by a small model (weak-to-strong generalization)

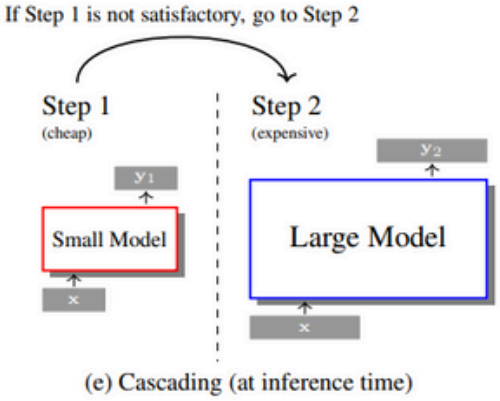
(b) Fine-tuning with KD Loss from a small model (weak-to-strong generalization)



(c) Data selection with a small model



(d) Ensemble of multiple small models



(e) Cascading (at inference time)

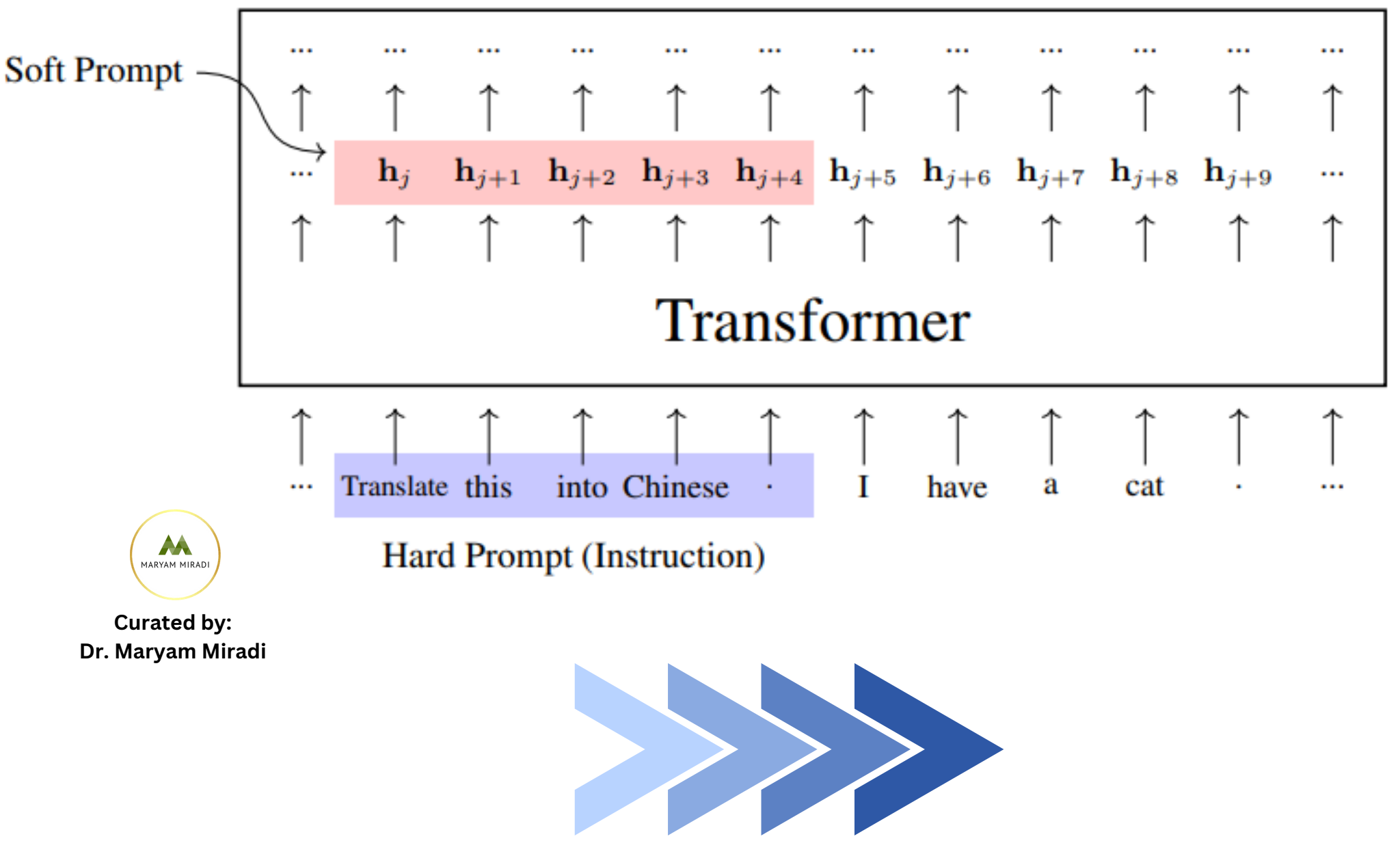
MARYAM MIRADI

Curated by:
Dr. Maryam Miradi

LLM Foundations

16/16

Soft Prompts



MARYAM MIRADI

Curated by:
Dr. Maryam Miradi