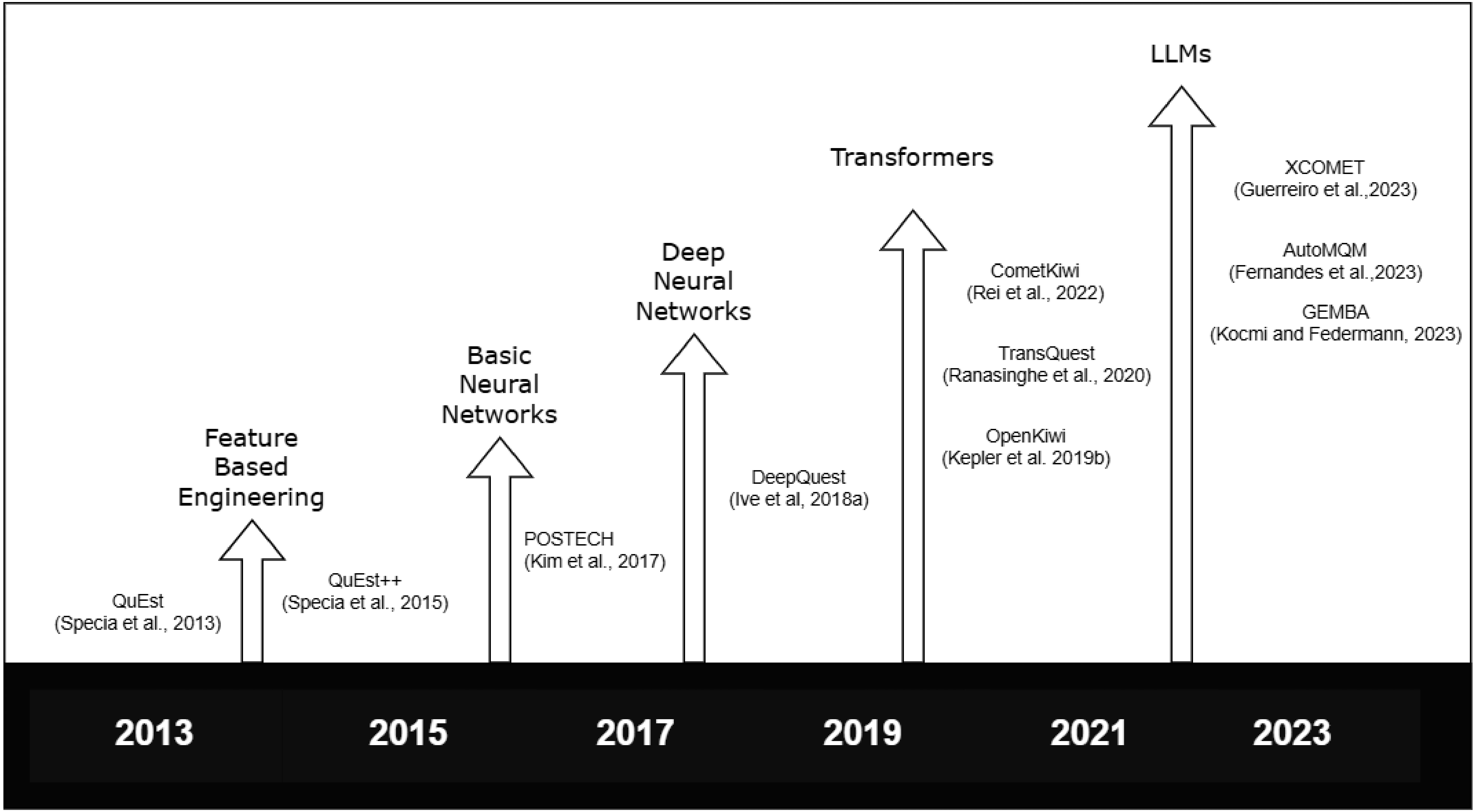
ALOPE: Adaptive Layer Optimization for Translation Quality Estimation using Large Language Models

Archchana Sindhujan, **Diptesh Kanojia**, Constantin Orasan,
Shenbin Quan , Cha Chu Chin Matthew

People-Centred AI
UNIVERSITY OF SURREY

Introduction

- ***Evaluating the quality*** of the translation is considered a challenging cross-lingual task - to ***identify the reliability*** of the translation and towards improvement of the translation systems.
- Traditional Machine Translation evaluation methods are Reference-based which are resource-intensive.
- **Quality estimation (QE)** aims to **estimate the quality of a translated text without the need for a reference translation** - Reduces cost and effort.
  - At different granularity levels – Segment-level, Word-level, Error Span Annotation, MQM


LLMs
Transformers
XCOMET
(Guerreiro et al.,2023)
AutoMQM
(Fernandes et al.,2023)
GEMBA
(Kocmi and Federmann, 2023)
CometKiwi
(Rei et al., 2022)
TransQuest
(Ranasinghe et al., 2020)
OpenKiwi
(Kepler et al. 2019b)
Deep
Neural
Networks
DeepQuest
(Ive et al, 2018a)
Basic
Neural
Networks
POSTECH
(Kim et al., 2017)
Feature
Based
Engineering
QuEst++
(Specia et al., 2015)
QuEst
(Specia et al., 2013)
2013
2015
2017
2019
2021
2023

Quality Estimation @ Sentence/Segment-level

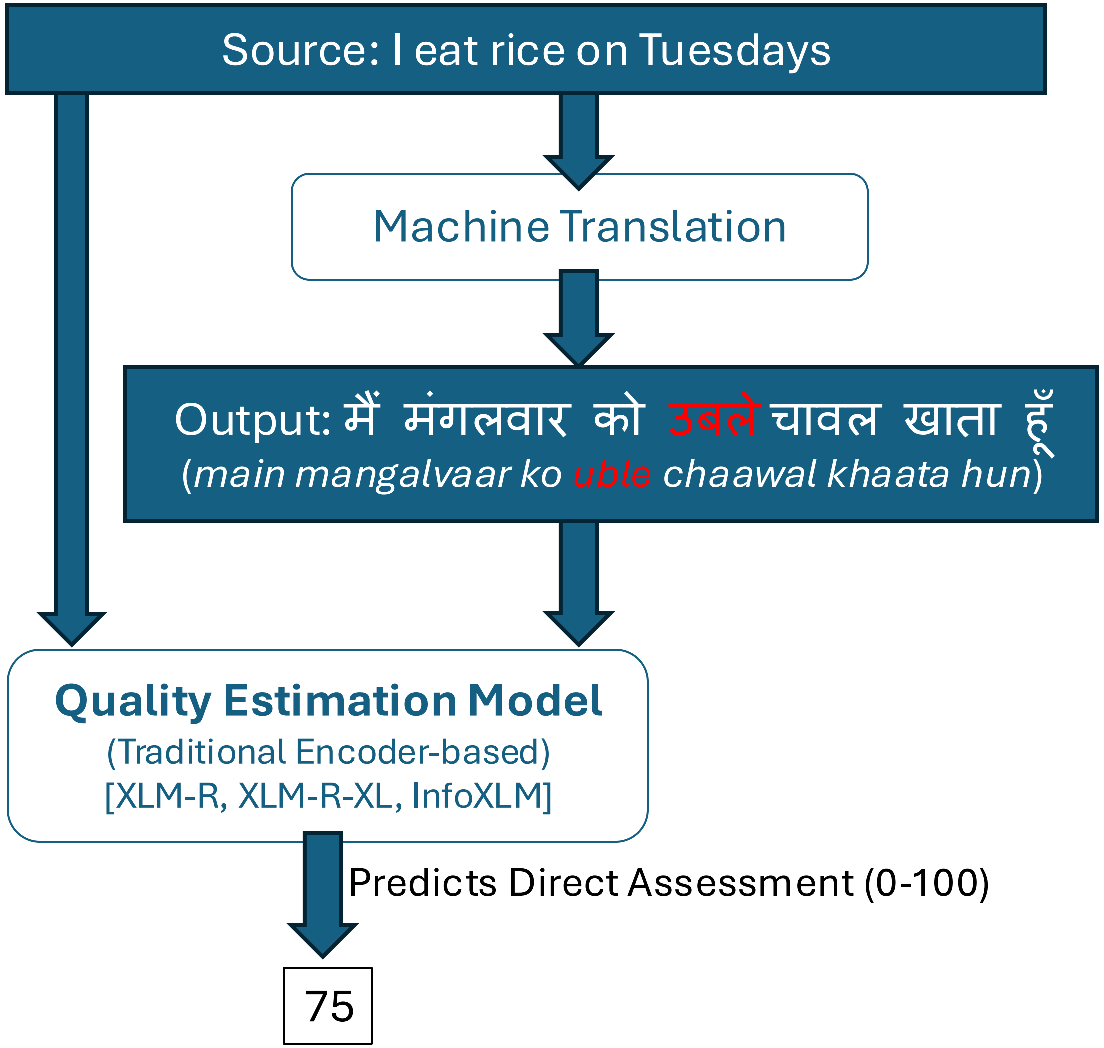
To provide a reliable, automatic measure of translation quality, crucial for system development and user-facing applications – without using a reference.

- Metrics like BLEU, chrF, MetricX **need a reference**.
- MT is subjective – multiple references – free order.

Quality Estimation (QE) is task of assessing the quality of machine-translated text in the absence of a human reference translation.

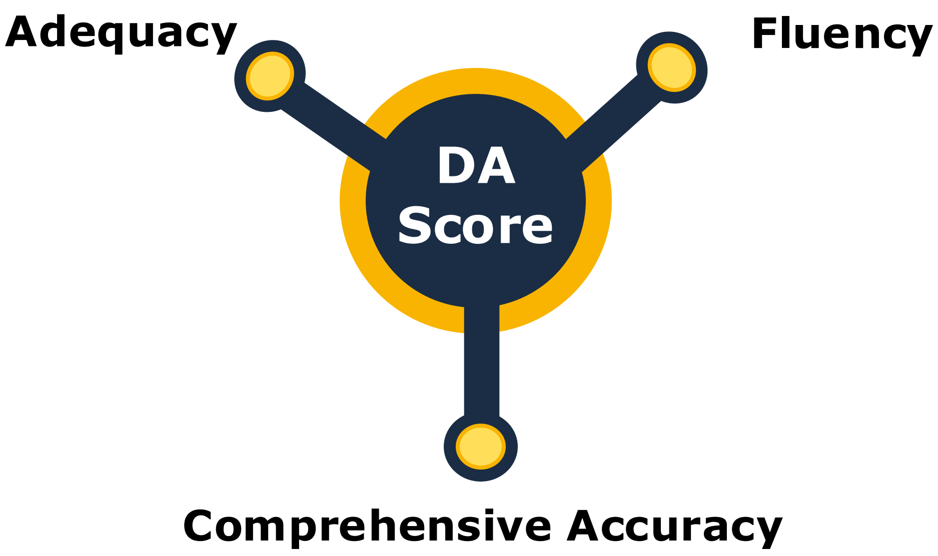
- **Segment-Level QE** focuses on assigning a quality score to a translated sentence, typically a Direct Assessment (DA) score from 0-100.

- **Word-Level QE** focuses on tagging each token in source and MT output with a OK/BAD tag, given the translation errors.



Direct Assessment (DA) score

Direct Assessment evaluates machine-translated content by having human assessors rate its quality on a scale of 0 to 100.



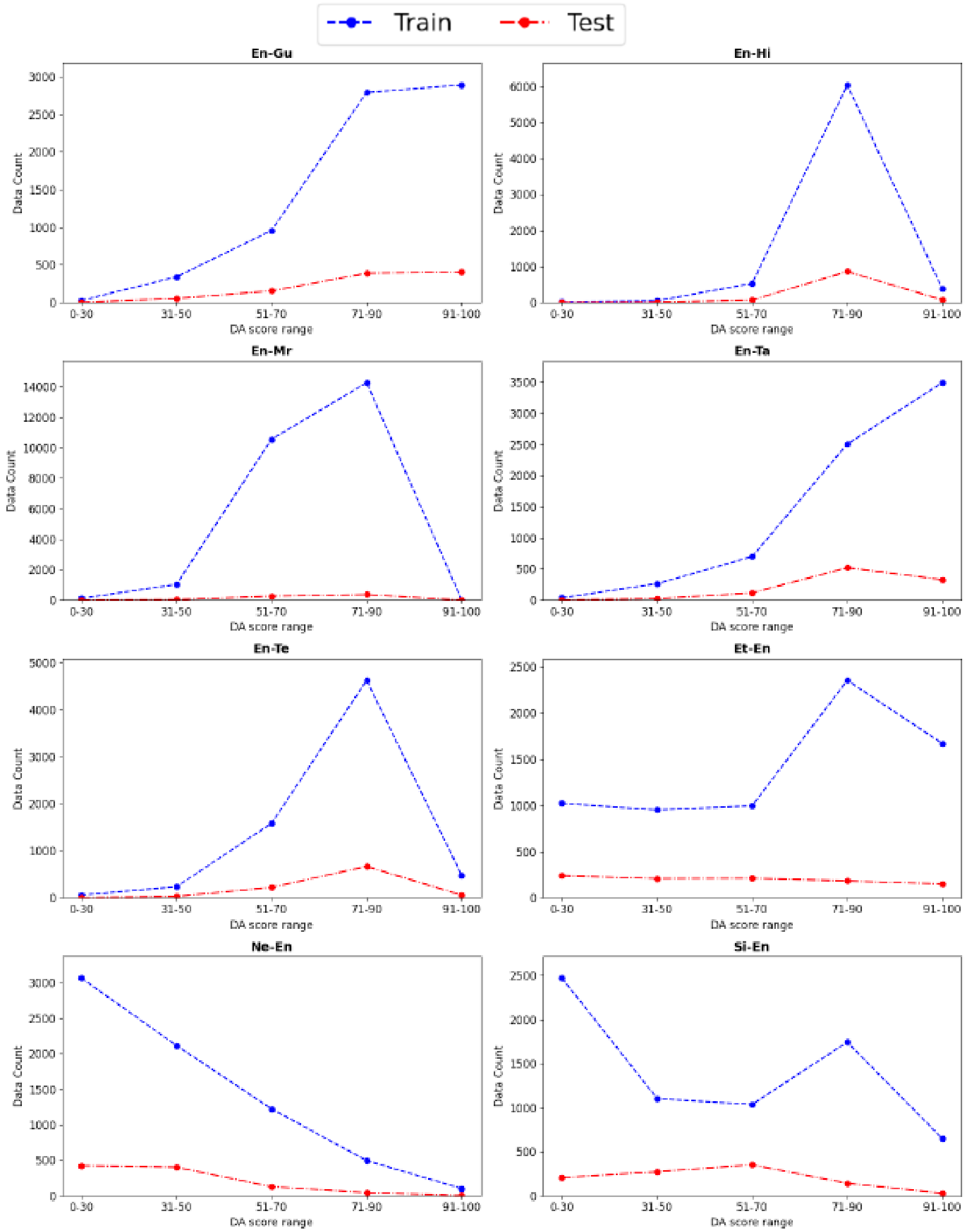
Overall Score	Translation conveys source meaning?	How much translation conveys to source?
1 - 10	**Completely inaccurate.**	The MT output is unintelligible. Studying the meaning of the sentence is hopeless; even allowing for context, one feels that guessing would be too unreliable. • The translation is incomprehensible, and machine translated output contains a mix of languages/dialects [which are not in the target language] (Adequacy) • None of the keywords are translated in the target language. (Adequacy) • There are major grammatical errors and typos (Fluency)
11 - 30	**Inaccurate but contains some keywords.**	• Incomprehensible translation (Fluency) • Translation contains some keywords but not all (Adequacy) • There are numerous grammatical errors and typos.
31 -50	**Partially. Target reflects partially the source.**	The general idea of the MT output is intelligible only after considerable study. • Translation is only partially understandable, and the overall meaning is not conveyed. (Adequacy and Fluency) • Translation contains some keywords (Adequacy) • There are many grammatical errors and typos (Fluency)
51 - 70	**Yes. Target reflects the overall meaning.**	The MT output is generally clear and intelligible. Despite some inaccuracies or infelicities of the sentence, one can understand (almost) immediately what it means. • Translation is understandable and reflects the source meaning. (Fluency) • Translation contains most keywords (Adequacy) • Only minor grammar errors (Fluency)
71 - 90	**Yes. Target reflects the source meaning without errors.**	The MT output is perfectly clear and intelligible. It is grammatical and reads like ordinary text. • Translation is very closed to the source meaning (Fluency) • Translation contains all keywords (Adequacy) • No errors but there are better word choices in the target language. (Adequacy)
91 - 100	**Yes. Target reflects source meaning without errors.**	• Perfect translation (Adequacy and Fluency) • Accurately reflects the meaning of the source (Fluency) • No errors

Dataset

- Focus on low-resource language pairs for QE.
- We utilize the DA score dataset from WMT QE shared task for our experiments.

  - En-Gu : English to ગુજરાતી (Gujarati)
  - En-Hi : English to **हिन्दी** (Hindi)
  - En-Mr : English to **मराठी** (Marathi)
  - En-Ta : English to தமிழ் (Tamil)
  - En-Te : English to తెలుగు (Telugu)

  - Et-En : **eesti keel** (Estonian) to English
  - Ne-En : **नेपाली** (Nepali) to English
  - Si-En : සිංහල (Sinhala) to English



GEMBA (Kochmi et. al., 2023) Prompt

```
Score the following translation from {Source Language} to {Target Language} on a continuous scale from 0 to 100, where score of zero means "no meaning preserved" and score of one hundred means "perfect meaning and grammar".

{Source Language} source: {Source Sentence}

{Target Language} translation: {Translated Sentence}

Score:
```

Experimental Settings

- **Zero-Shot :** Model generates outputs using pre-trained knowledge and the generalization ability.

- **Standard Instruction-Tuning :** Adapting a model using a dataset that includes explicit instructions for specific tasks.

- **ALOPE :** Enhances fine-tuning by leveraging regression headers across multiple Transformer layers.

Models

- All experiments conducted with open-source LLMs with 8B parameters or less:
  - Llama-2-7B
  - Llama 3.1-8B
  - Llama 3.2-3B
  - Aya-expanse-8B

Challenges for LLM-based QE

- LLMs are **optimized for next-token prediction**, excelling at generative tasks but struggling with regression-based goals like quality estimation which require numerical and cross-lingual reasoning.
  - This limits their ability to capture fine-grained relationships between language features and numerical scores.

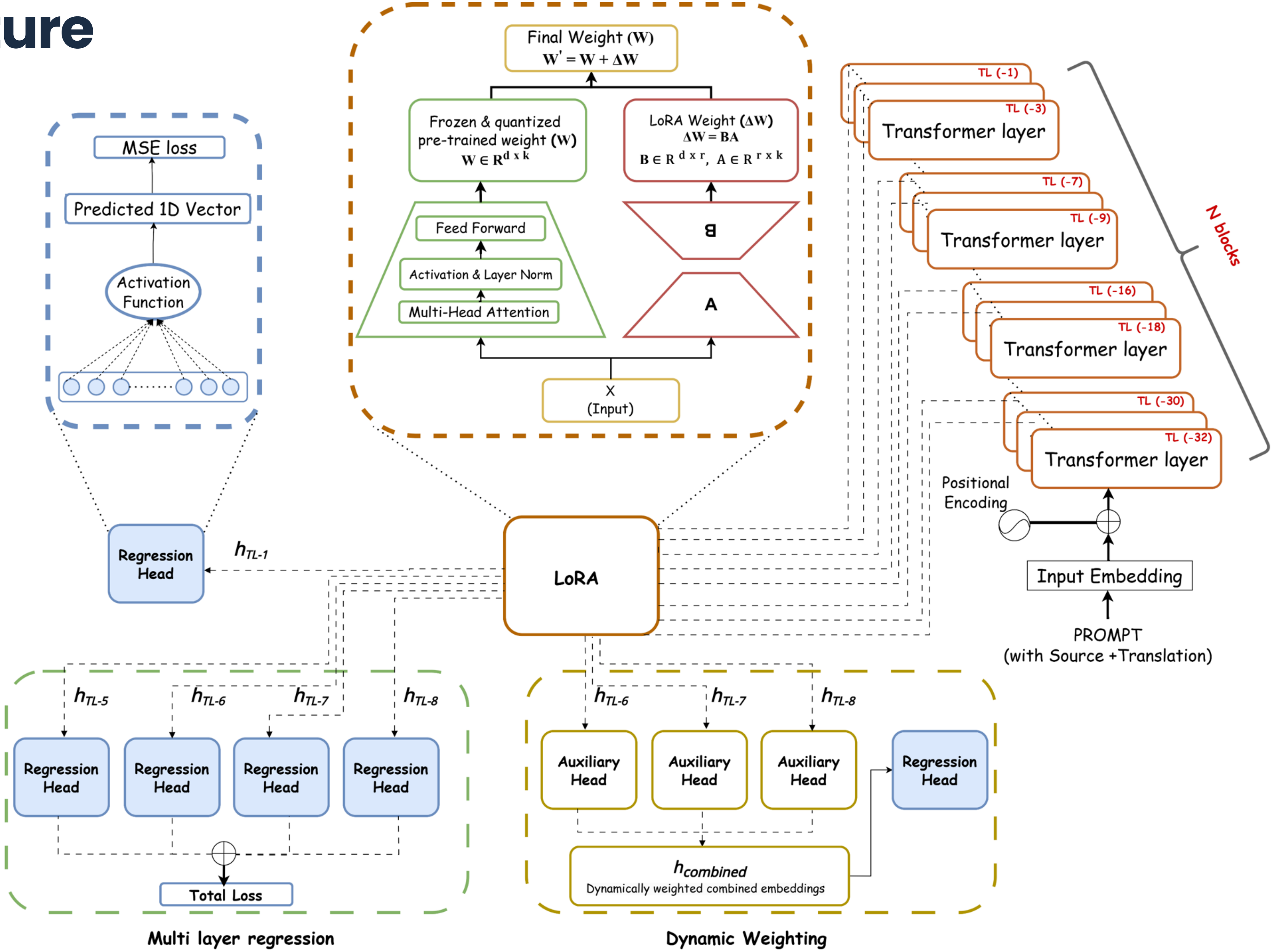
- Although LLMs refine context through multiple Transformer layers, standard practice usually <u>predicts with the final layer for output</u>.

Why ALOPE (Adaptive Layer Optimization)?

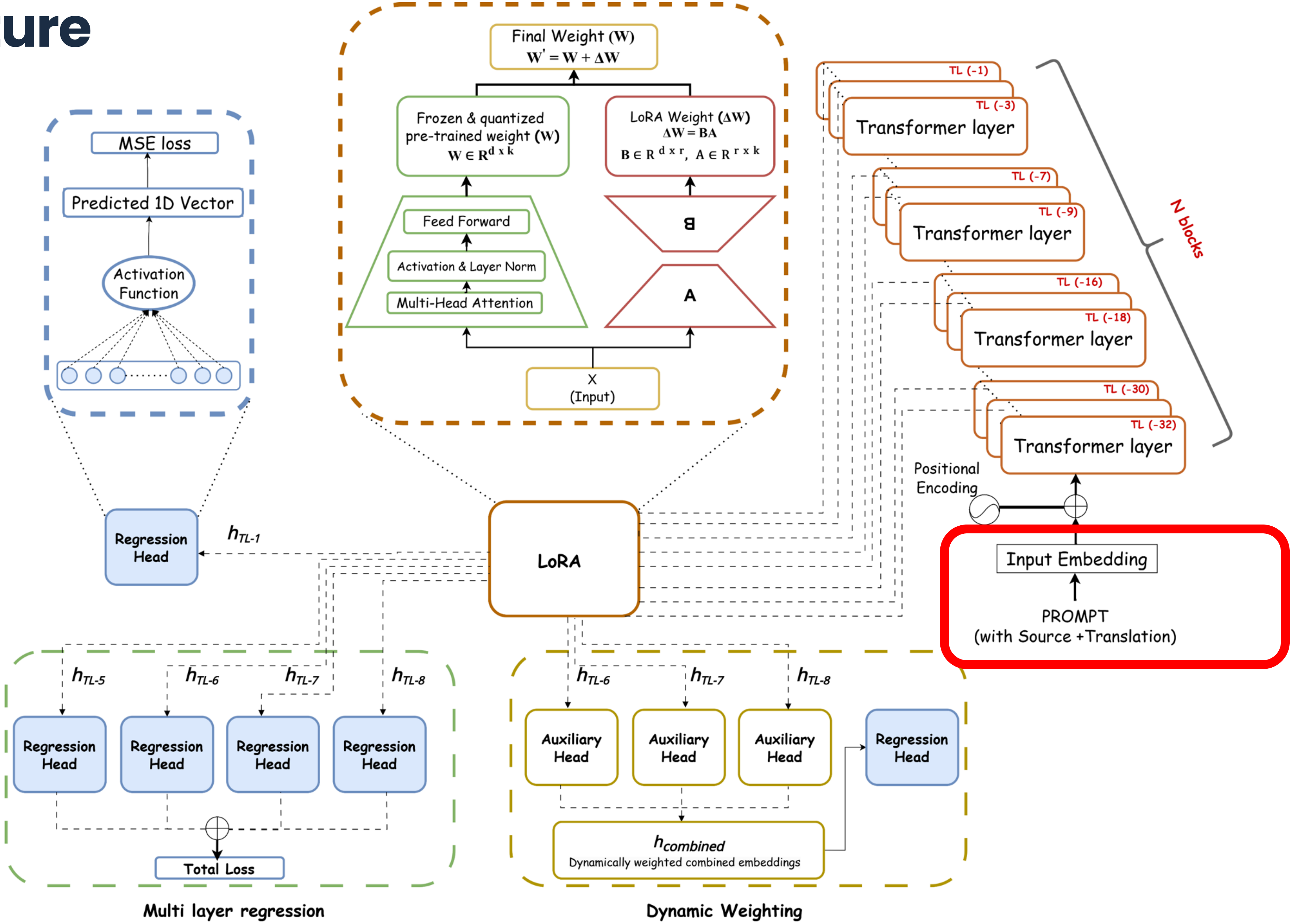
- ALOPE enables LLMs to perform regression by integrating a regression head within the low-rank adapted Transformer architecture.

- We analyze the impact of adaptive regression heads to determine which Transformer layers contribute most to enhancing the performance of LLMs for QE with low-resource language pairs.

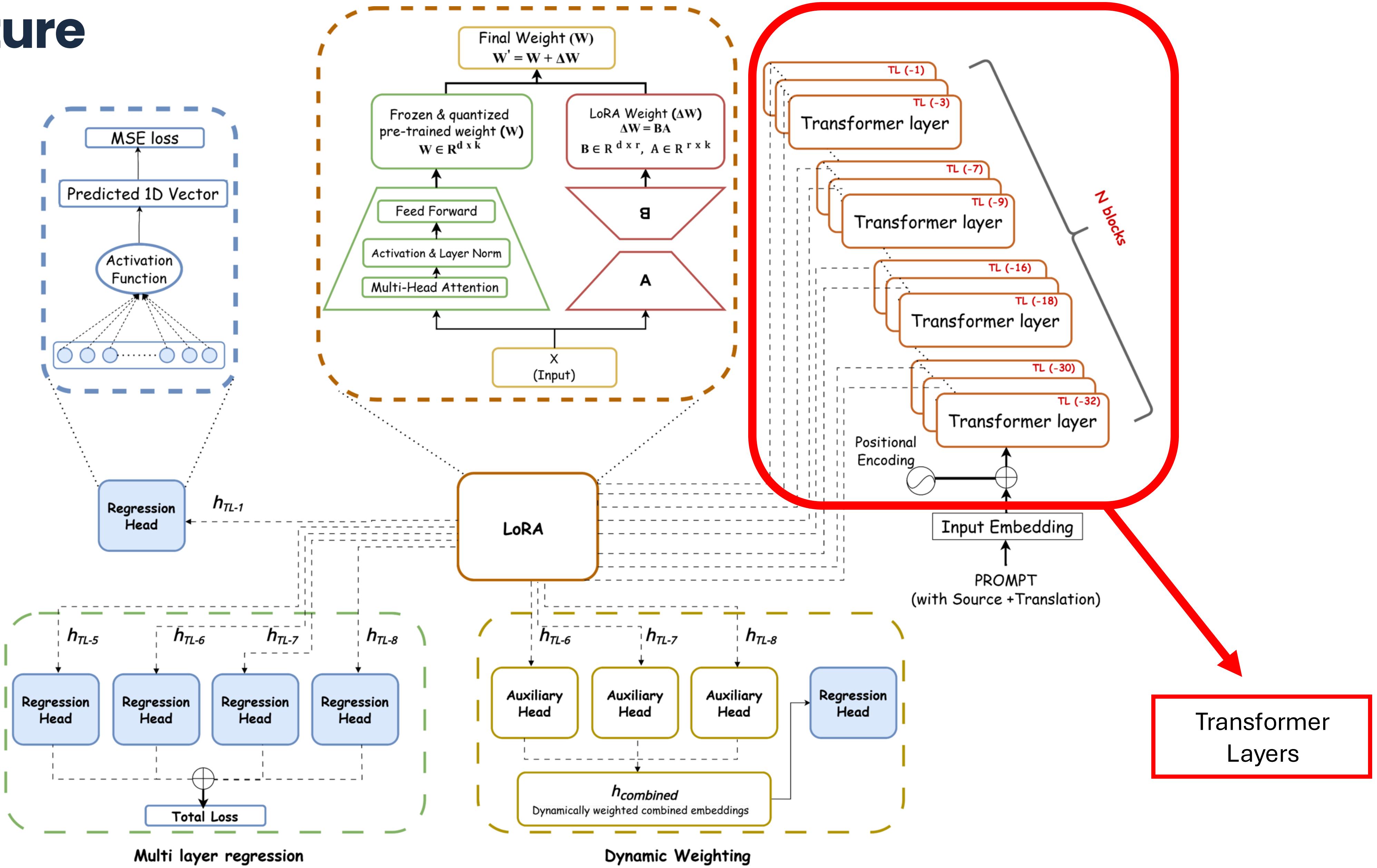
Architecture



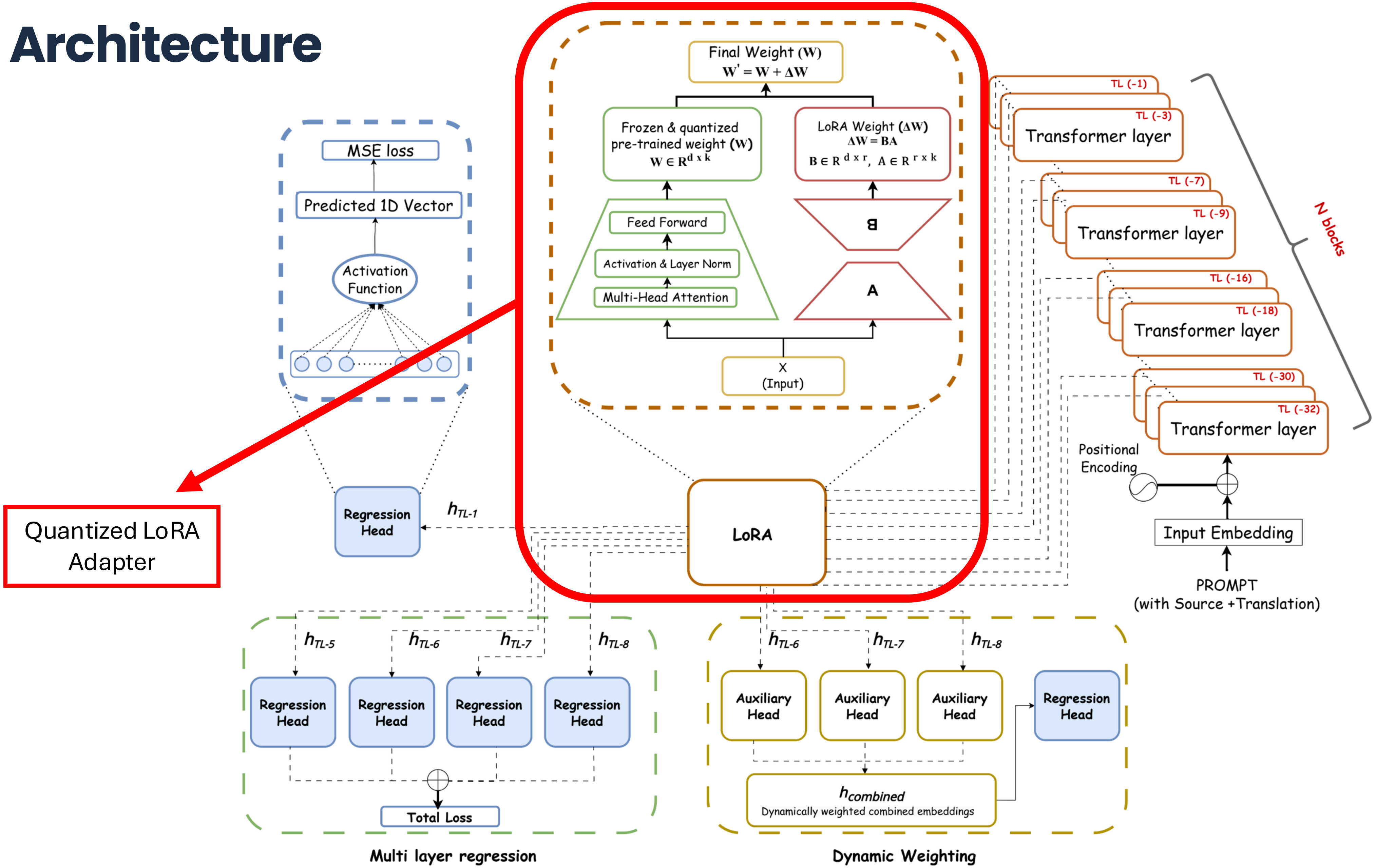
Architecture



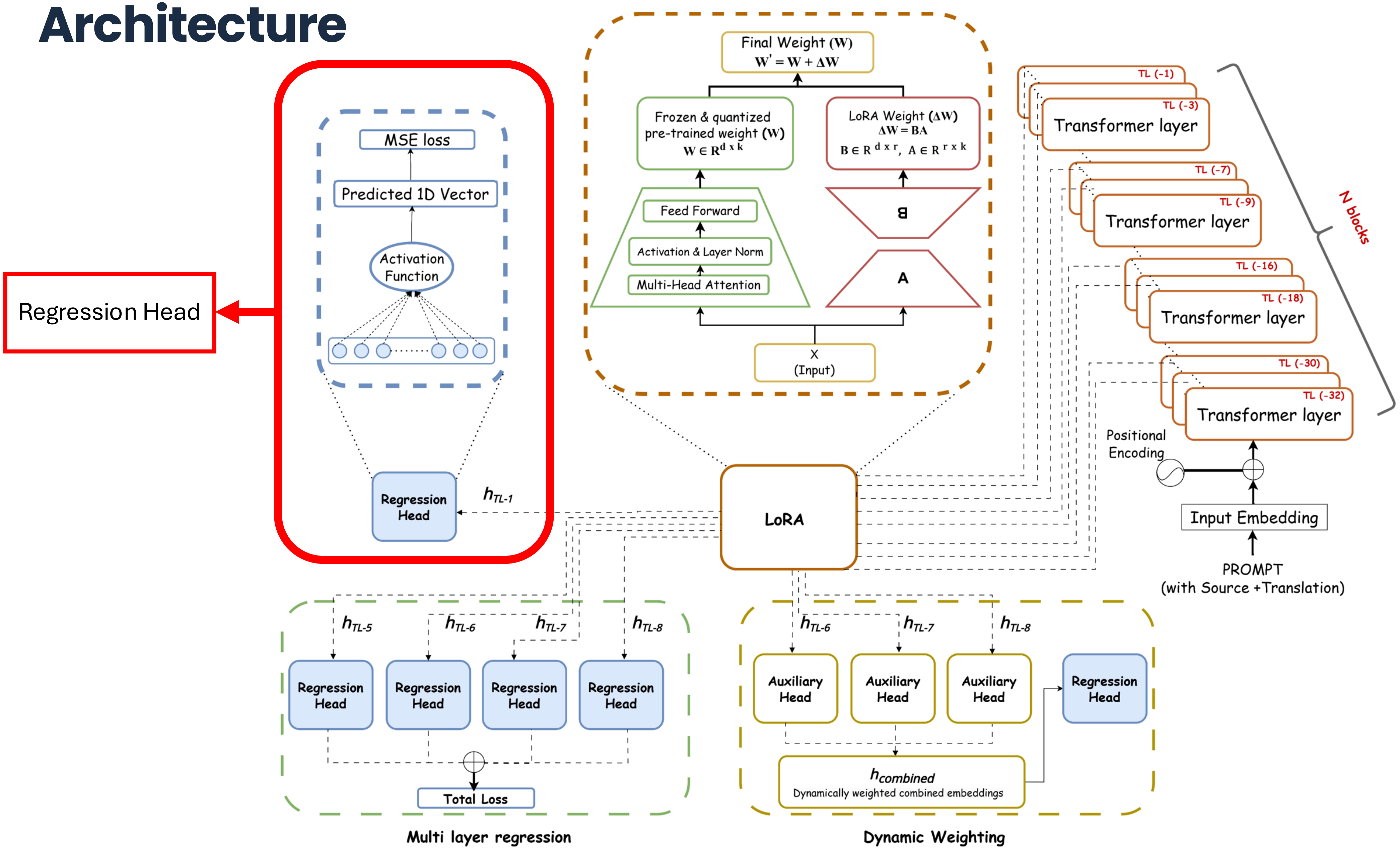
Architecture



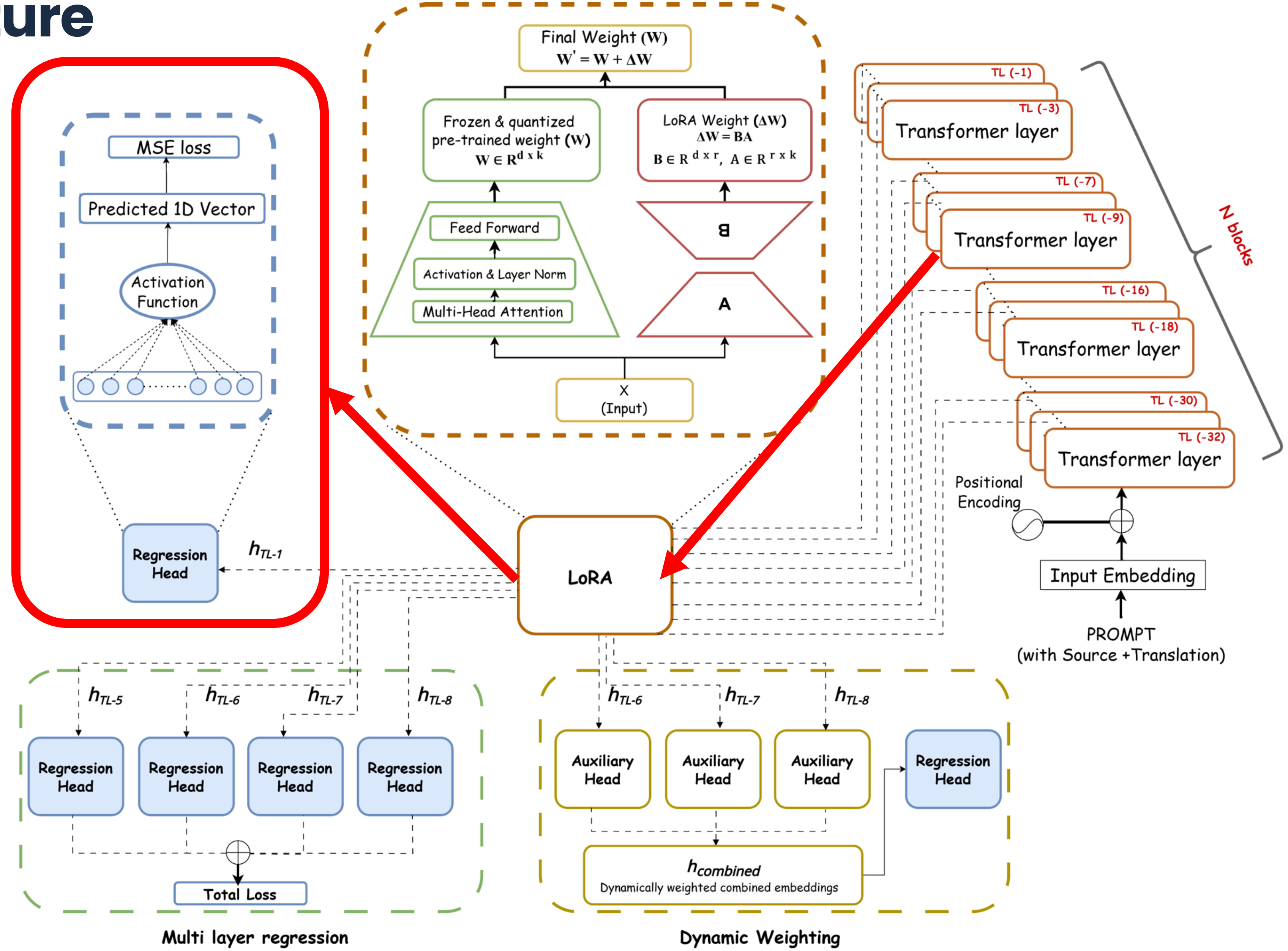
Architecture



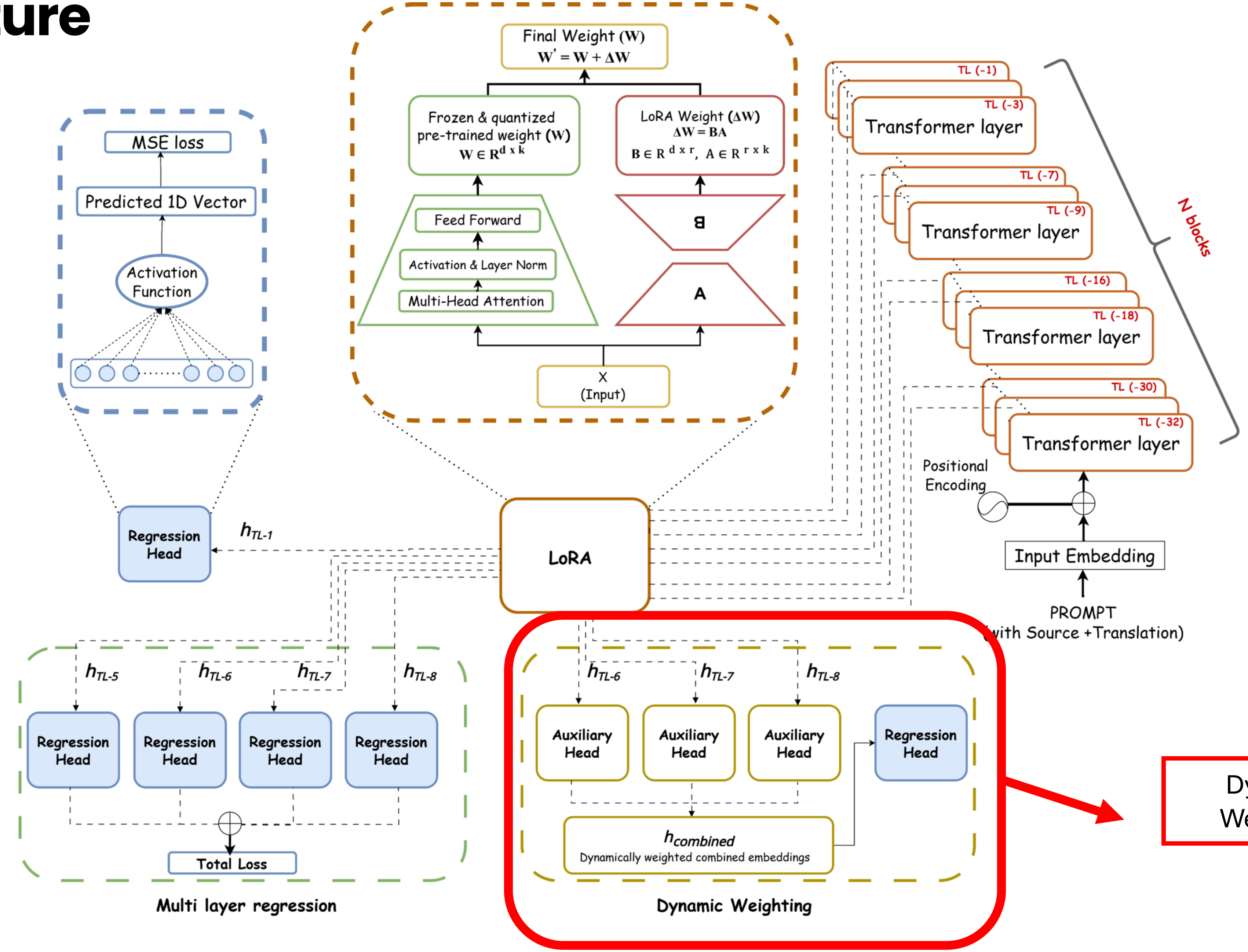
Architecture



Architecture

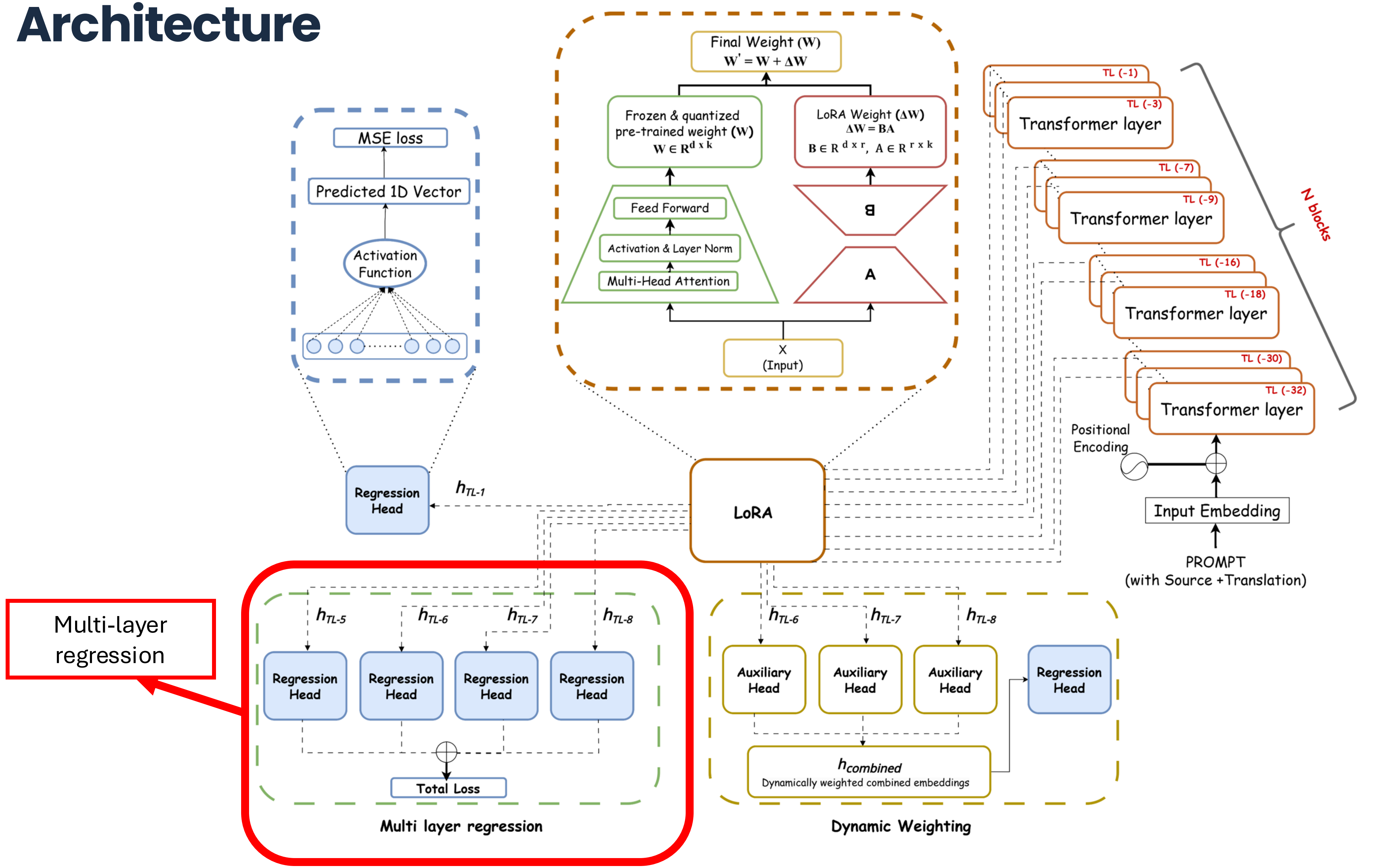


Architecture



Dynamic Weighting

Architecture



Evaluation Metrics for DA prediction in QE

- **Spearman's correlation** between predicted scores and human DA mean scores

- Applied **Williams test** to compare correlation significance between models and to identify <u>whether top models outperform others</u> **statistically.**

ALOPE

		Model	En-Gu	En-Hi	En-Mr	En-Ta	En-Te	Et-En	Ne-En	Si-En
Final Transformer Layer	TL (-1)	llama-2-7b	0.563	0.414	0.609	0.525	0.356	0.742	0.596	0.565
		llama 3.1-8B	0.594	0.469	0.620	0.567	0.363	0.734	0.647	0.547
		llama 3.2-3B	0.604	0.477	0.636	0.580	0.348	0.735	0.674	0.543
		aya-expanse-8b	0.068	0.178	0.219	-0.006	0.275	0.115	0.012	0.077
Intermediate Transformer Layer	TL (-7)	llama-2-7b	0.567	0.336	0.542	0.484	0.317	0.739	0.606	0.573
		llama 3.1-8B	0.590	0.477	0.625	0.528	0.388	0.744	0.638	0.544
		llama 3.2-3B	0.606	0.479	0.617	0.585	0.369	0.751	0.664	0.553
		aya-expanse-8b	0.538	0.447	0.597	0.528	0.347	0.741	0.646	0.544
Intermediate Transformer Layer	TL (-11)	llama-2-7b	0.360	0.301	0.361	0.254	0.293	0.405	0.164	0.049
		llama 3.1-8B	0.514	0.412	0.609	0.438	0.304	0.148	0.554	0.493
		llama 3.2-3B	0.594	0.476	0.605	0.610	0.373	0.748	0.678	0.560
		aya-expanse-8b	0.490	0.411	0.572	0.445	0.336	0.569	0.453	0.439
Intermediate Transformer Layer	TL (-16)	llama-2-7b	0.540	0.381	0.585	0.482	0.308	0.751	0.580	0.569
		llama 3.1-8B	0.558	0.453	0.602	0.523	0.350	0.737	0.652	0.513
		llama 3.2-3B	0.557	0.459	0.597	0.547	0.338	0.745	0.682	0.567
		aya-expanse-8b	0.467	0.390	0.557	0.481	0.314	0.727	0.576	0.540
Lower-level Transformer Layer	TL (-20)	llama-2-7b	0.470	0.405	0.544	0.460	0.338	0.684	0.508	0.534
		llama 3.1-8B	0.484	0.394	0.553	0.321	0.172	0.649	0.524	0.494
		llama 3.2-3B	0.430	0.408	0.579	0.303	0.286	0.601	0.488	0.464
		aya-expanse-8b	0.437	0.300	0.488	0.263	0.287	0.483	0.438	0.395
Lower-level Transformer Layer	TL (-24)	llama-2-7b	0.500	0.421	0.538	0.379	0.239	0.630	0.507	0.472
		llama 3.1-8B	0.421	0.378	0.552	0.330	0.290	0.515	0.530	0.464
		llama 3.2-3B	0.443	0.376	0.507	0.367	0.299	0.559	0.528	0.487
		aya-expanse-8b	0.375	0.319	0.440	0.337	0.220	0.393	0.407	0.345

ALOPE

	Model	En-Gu	En-Hi	En-Mr	En-Ta	En-Te	Et-En	Ne-En	Si-En
TL (-1)	llama-2-7b	0.563	0.414	0.609	0.525	0.356	0.742	0.596	0.565
	llama 3.1-8B	0.594	0.469	0.620	0.567	0.363	0.734	0.647	0.547
	llama 3.2-3B	0.604	0.477	**0.636**	0.580	0.348	0.735	0.674	0.543
	aya-expanse-8b	0.068	0.178	0.219	-0.006	0.275	0.115	0.012	0.077
TL (-7)	llama-2-7b	0.567	0.336	0.542	0.484	0.317	0.739	0.606	**0.573**
	llama 3.1-8B	0.590	0.477	0.625	0.528	**0.388**	0.744	0.638	0.544
	llama 3.2-3B	**0.606**	**0.479**	0.617	0.585	0.369	**0.751**	0.664	0.553
	aya-expanse-8b	0.538	0.447	0.597	0.528	0.347	0.741	0.646	0.544
TL (-11)	llama-2-7b	0.360	0.301	0.361	0.254	0.293	0.405	0.164	0.049
	llama 3.1-8B	0.514	0.412	0.609	0.438	0.304	0.148	0.554	0.493
	llama 3.2-3B	0.594	0.476	0.605	**0.610**	0.373	0.748	0.678	0.560
	aya-expanse-8b	0.490	0.411	0.572	0.445	0.336	0.569	0.453	0.439
TL (-16)	llama-2-7b	0.540	0.381	0.585	0.482	0.308	0.751	0.580	0.569
	llama 3.1-8B	0.558	0.453	0.602	0.523	0.350	0.737	0.652	0.513
	llama 3.2-3B	0.557	0.459	0.597	0.547	0.338	0.745	**0.682**	0.567
	aya-expanse-8b	0.467	0.390	0.557	0.481	0.314	0.727	0.576	0.540
TL (-20)	llama-2-7b	0.470	0.405	0.544	0.460	0.338	0.684	0.508	0.534
	llama 3.1-8B	0.484	0.394	0.553	0.321	0.172	0.649	0.524	0.494
	llama 3.2-3B	0.430	0.408	0.579	0.303	0.286	0.601	0.488	0.464
	aya-expanse-8b	0.437	0.300	0.488	0.263	0.287	0.483	0.438	0.395
TL (-24)	llama-2-7b	0.500	0.421	0.538	0.379	0.239	0.630	0.507	0.472
	llama 3.1-8B	0.421	0.378	0.552	0.330	0.290	0.515	0.530	0.464
	llama 3.2-3B	0.443	0.376	0.507	0.367	0.299	0.559	0.528	0.487
	aya-expanse-8b	0.375	0.319	0.440	0.337	0.220	0.393	0.407	0.345

ALOPE

	Model	En-Gu	En-Hi	En-Mr	En-Ta	En-Te	Et-En	Ne-En	Si-En	*Avg.*
TL (-1)	llama-2-7b	0.563	0.414	0.609	0.525	0.356	0.742	0.596	0.565	*0.546*
	llama 3.1-8B	0.594	0.469	0.620	0.567	0.363	0.734	0.647	0.547	*0.567*
	llama 3.2-3B	0.604	0.477	**0.636**	0.580	0.348	0.735	0.674	0.543	*0.575*
	aya-expanse-8b	0.068	0.178	0.219	-0.006	0.275	0.115	0.012	0.077	*0.117*
	Avg	*0.457*	*0.385*	*0.521*	*0.416*	*0.335*	*0.581*	*0.482*	*0.433*	
TL (-7)	llama-2-7b	0.567	0.336	0.542	0.484	0.317	0.739	0.606	**0.573**	*0.520*
	llama 3.1-8B	0.590	0.477	0.625	0.528	**0.388**	0.744	0.638	0.544	*0.567*
	llama 3.2-3B	**0.606**	**0.479**	0.617	0.585	0.369	**0.751**	0.664	0.553	*0.578*
	aya-expanse-8b	0.538	0.447	0.597	0.528	0.347	0.741	0.646	0.544	*0.549*
	Avg	*0.575*	*0.435*	*0.595*	*0.531*	*0.355*	*0.744*	*0.639*	*0.554*	
TL (-11)	llama-2-7b	0.360	0.301	0.361	0.254	0.293	0.405	0.164	0.049	*0.273*
	llama 3.1-8B	0.514	0.412	0.609	0.438	0.304	0.148	0.554	0.493	*0.434*
	llama 3.2-3B	0.594	0.476	0.605	**0.610**	0.373	0.748	0.678	0.560	*0.581*
	aya-expanse-8b	0.490	0.411	0.572	0.445	0.336	0.569	0.453	0.439	*0.464*
	Avg	*0.489*	*0.400*	*0.537*	*0.437*	*0.327*	*0.467*	*0.462*	*0.385*	
TL (-16)	llama-2-7b	0.540	0.381	0.585	0.482	0.308	0.751	0.580	0.569	*0.524*
	llama 3.1-8B	0.558	0.453	0.602	0.523	0.350	0.737	0.652	0.513	*0.548*
	llama 3.2-3B	0.557	0.459	0.597	0.547	0.338	0.745	**0.682**	0.567	*0.561*
	aya-expanse-8b	0.467	0.390	0.557	0.481	0.314	0.727	0.576	0.540	*0.506*
	Avg	*0.530*	*0.421*	*0.585*	*0.508*	*0.327*	*0.740*	*0.622*	*0.547*	
TL (-20)	llama-2-7b	0.470	0.405	0.544	0.460	0.338	0.684	0.508	0.534	*0.493*
	llama 3.1-8B	0.484	0.394	0.553	0.321	0.172	0.649	0.524	0.494	*0.449*
	llama 3.2-3B	0.430	0.408	0.579	0.303	0.286	0.601	0.488	0.464	*0.445*
	aya-expanse-8b	0.437	0.300	0.488	0.263	0.287	0.483	0.438	0.395	*0.386*
	Avg	*0.455*	*0.377*	*0.541*	*0.337*	*0.271*	*0.604*	*0.490*	*0.472*	
TL (-24)	llama-2-7b	0.500	0.421	0.538	0.379	0.239	0.630	0.507	0.472	*0.461*
	llama 3.1-8B	0.421	0.378	0.552	0.330	0.290	0.515	0.530	0.464	*0.435*
	llama 3.2-3B	0.443	0.376	0.507	0.367	0.299	0.559	0.528	0.487	*0.446*
	aya-expanse-8b	0.375	0.319	0.440	0.337	0.220	0.393	0.407	0.345	*0.354*
	Avg	0.435	0.373	0.509	0.353	0.262	0.524	0.493	0.442	

ALOPE

	Model	En-Gu	En-Hi	En-Mr	En-Ta	En-Te	Et-En	Ne-En	Si-En	*Avg.*
TL (-1)	llama-2-7b	0.563	0.414	0.609	0.525	0.356	0.742	0.596	0.565	*0.546*
	llama 3.1-8B	0.594	0.469	0.620	0.567	0.363	0.734	0.647	0.547	*0.567*
	llama 3.2-3B	0.604	0.477	**0.636**	0.580	0.348	0.735	0.674	0.543	*0.575*
	aya-expanse-8b	0.068	0.178	0.219	-0.006	0.275	0.115	0.012	0.077	*0.117*
	Avg	*0.457*	*0.385*	*0.521*	*0.416*	*0.335*	*0.581*	*0.482*	*0.433*	
TL (-7)	llama-2-7b	0.567	0.336	0.542	0.484	0.317	0.739	0.606	**0.573**	*0.520*
	llama 3.1-8B	0.590	0.477	0.625	0.528	**0.388**	0.744	0.638	0.544	*0.567*
	llama 3.2-3B	**0.606**	**0.479**	0.617	0.585	0.369	**0.751**	0.664	0.553	*0.578*
	aya-expanse-8b	0.538	0.447	0.597	0.528	0.347	0.741	0.646	0.544	*0.549*
	Avg	*0.575*	*0.435*	*0.595*	*0.531*	*0.355*	*0.744*	*0.639*	*0.554*	
TL (-11)	llama-2-7b	0.360	0.301	0.361	0.254	0.293	0.405	0.164	0.049	*0.273*
	llama 3.1-8B	0.514	0.412	0.609	0.438	0.304	0.148	0.554	0.493	*0.434*
	llama 3.2-3B	0.594	0.476	0.605	**0.610**	0.373	0.748	0.678	0.560	*0.581*
	aya-expanse-8b	0.490	0.411	0.572	0.445	0.336	0.569	0.453	0.439	*0.464*
	Avg	*0.489*	*0.400*	*0.537*	*0.437*	*0.327*	*0.467*	*0.462*	*0.385*	
TL (-16)	llama-2-7b	0.540	0.381	0.585	0.482	0.308	0.751	0.580	0.569	*0.524*
	llama 3.1-8B	0.558	0.453	0.602	0.523	0.350	0.737	0.652	0.513	*0.548*
	llama 3.2-3B	0.557	0.459	0.597	0.547	0.338	0.745	**0.682**	0.567	*0.561*
	aya-expanse-8b	0.467	0.390	0.557	0.481	0.314	0.727	0.576	0.540	*0.506*
	Avg	*0.530*	*0.421*	*0.585*	*0.508*	*0.327*	*0.740*	*0.622*	*0.547*	
TL (-20)	llama-2-7b	0.470	0.405	0.544	0.460	0.338	0.684	0.508	0.534	*0.493*
	llama 3.1-8B	0.484	0.394	0.553	0.321	0.172	0.649	0.524	0.494	*0.449*
	llama 3.2-3B	0.430	0.408	0.579	0.303	0.286	0.601	0.488	0.464	*0.445*
	aya-expanse-8b	0.437	0.300	0.488	0.263	0.287	0.483	0.438	0.395	*0.386*
	Avg	*0.455*	*0.377*	*0.541*	*0.337*	*0.271*	*0.604*	*0.490*	*0.472*	
TL (-24)	llama-2-7b	0.500	0.421	0.538	0.379	0.239	0.630	0.507	0.472	*0.461*
	llama 3.1-8B	0.421	0.378	0.552	0.330	0.290	0.515	0.530	0.464	*0.435*
	llama 3.2-3B	0.443	0.376	0.507	0.367	0.299	0.559	0.528	0.487	*0.446*
	aya-expanse-8b	0.375	0.319	0.440	0.337	0.220	0.393	0.407	0.345	*0.354*
	Avg	0.435	0.373	0.509	0.353	0.262	0.524	0.493	0.442	

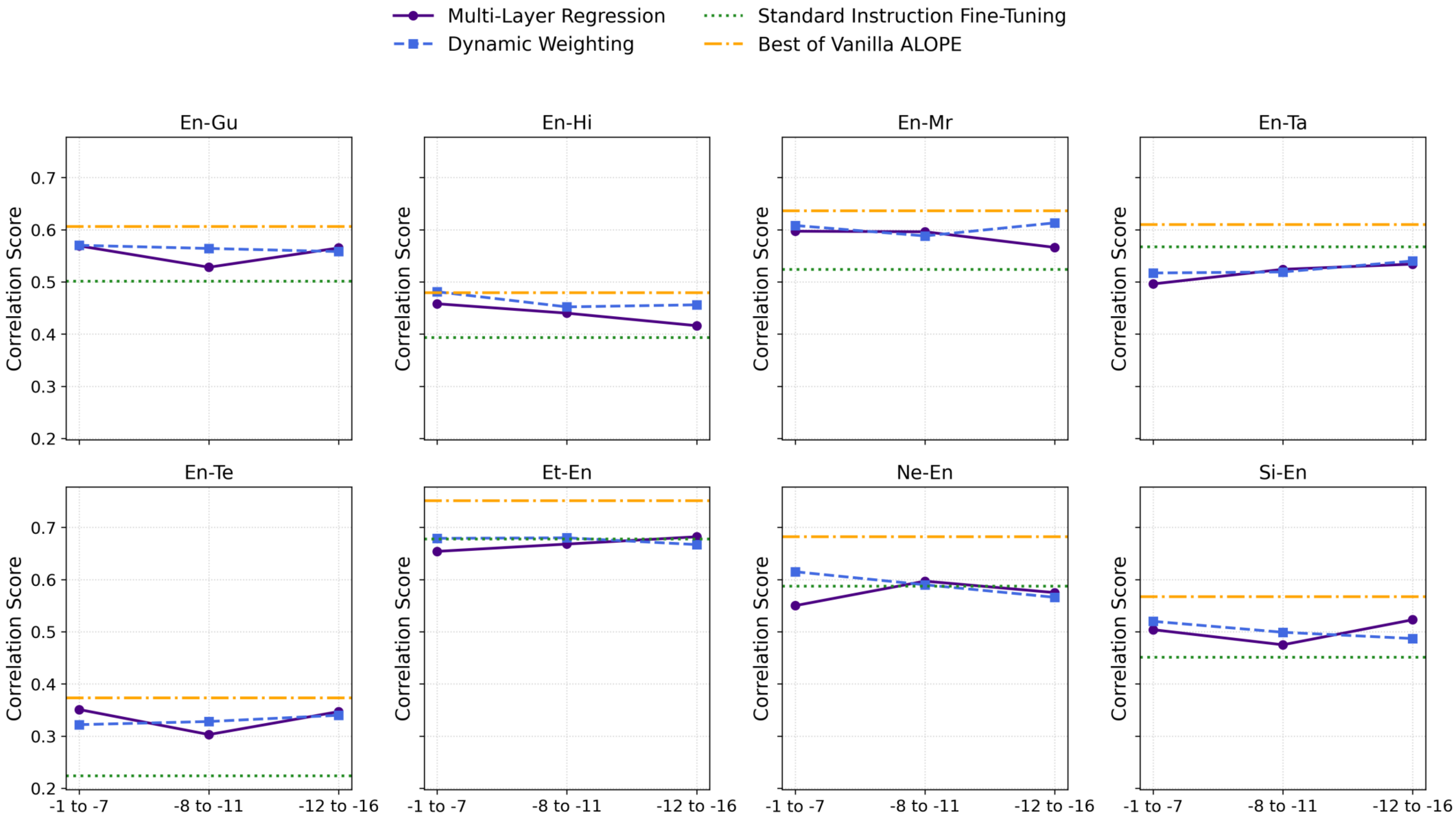
ALOPE

	Model	En-Gu	En-Hi	En-Mr	En-Ta	En-Te	Et-En	Ne-En	Si-En	*Avg.*
TL (-1)	llama-2-7b	0.563	0.414	0.609	0.525	0.356	0.742	0.596	0.565	*0.546*
	llama 3.1-8B	0.594	0.469	0.620	0.567	0.363	0.734	0.647	0.547	*0.567*
	llama 3.2-3B	0.604	0.477	**0.636**	0.580	0.348	0.735	0.674	0.543	*0.575*
	aya-expanse-8b	0.068	0.178	0.219	-0.006	0.275	0.115	0.012	0.077	*0.117*
	Avg	*0.457*	*0.385*	*0.521*	*0.416*	*0.335*	*0.581*	*0.482*	*0.433*	
TL (-7)	llama-2-7b	0.567	0.336	0.542	0.484	0.317	0.739	0.606	**0.573**	*0.520*
	llama 3.1-8B	0.590	0.477	0.625	0.528	**0.388**	0.744	0.638	0.544	*0.567*
	llama 3.2-3B	**0.606**	**0.479**	0.617	0.585	0.369	**0.751**	0.664	0.553	*0.578*
	aya-expanse-8b	0.538	0.447	0.597	0.528	0.347	0.741	0.646	0.544	*0.549*
	Avg	*0.575*	*0.435*	*0.595*	*0.531*	*0.355*	*0.744*	*0.639*	*0.554*	
TL (-11)	llama-2-7b	0.360	0.301	0.361	0.254	0.293	0.405	0.164	0.049	*0.273*
	llama 3.1-8B	0.514	0.412	0.609	0.438	0.304	0.148	0.554	0.493	*0.434*
	llama 3.2-3B	0.594	0.476	0.605	**0.610**	0.373	0.748	0.678	0.560	*0.581*
	aya-expanse-8b	0.490	0.411	0.572	0.445	0.336	0.569	0.453	0.439	*0.464*
	Avg	*0.489*	*0.400*	*0.537*	*0.437*	*0.327*	*0.467*	*0.462*	*0.385*	
TL (-16)	llama-2-7b	0.540	0.381	0.585	0.482	0.308	0.751	0.580	0.569	*0.524*
	llama 3.1-8B	0.558	0.453	0.602	0.523	0.350	0.737	0.652	0.513	*0.548*
	llama 3.2-3B	0.557	0.459	0.597	0.547	0.338	0.745	**0.682**	0.567	*0.561*
	aya-expanse-8b	0.467	0.390	0.557	0.481	0.314	0.727	0.576	0.540	*0.506*
	Avg	*0.530*	*0.421*	*0.585*	*0.508*	*0.327*	*0.740*	*0.622*	*0.547*	
TL (-20)	llama-2-7b	0.470	0.405	0.544	0.460	0.338	0.684	0.508	0.534	*0.493*
	llama 3.1-8B	0.484	0.394	0.553	0.321	0.172	0.649	0.524	0.494	*0.449*
	llama 3.2-3B	0.430	0.408	0.579	0.303	0.286	0.601	0.488	0.464	*0.445*
	aya-expanse-8b	0.437	0.300	0.488	0.263	0.287	0.483	0.438	0.395	*0.386*
	Avg	*0.455*	*0.377*	*0.541*	*0.337*	*0.271*	*0.604*	*0.490*	*0.472*	
TL (-24)	llama-2-7b	0.500	0.421	0.538	0.379	0.239	0.630	0.507	0.472	*0.461*
	llama 3.1-8B	0.421	0.378	0.552	0.330	0.290	0.515	0.530	0.464	*0.435*
	llama 3.2-3B	0.443	0.376	0.507	0.367	0.299	0.559	0.528	0.487	*0.446*
	aya-expanse-8b	0.375	0.319	0.440	0.337	0.220	0.393	0.407	0.345	*0.354*
	Avg	0.435	0.373	0.509	0.353	0.262	0.524	0.493	0.442	

ALOPE

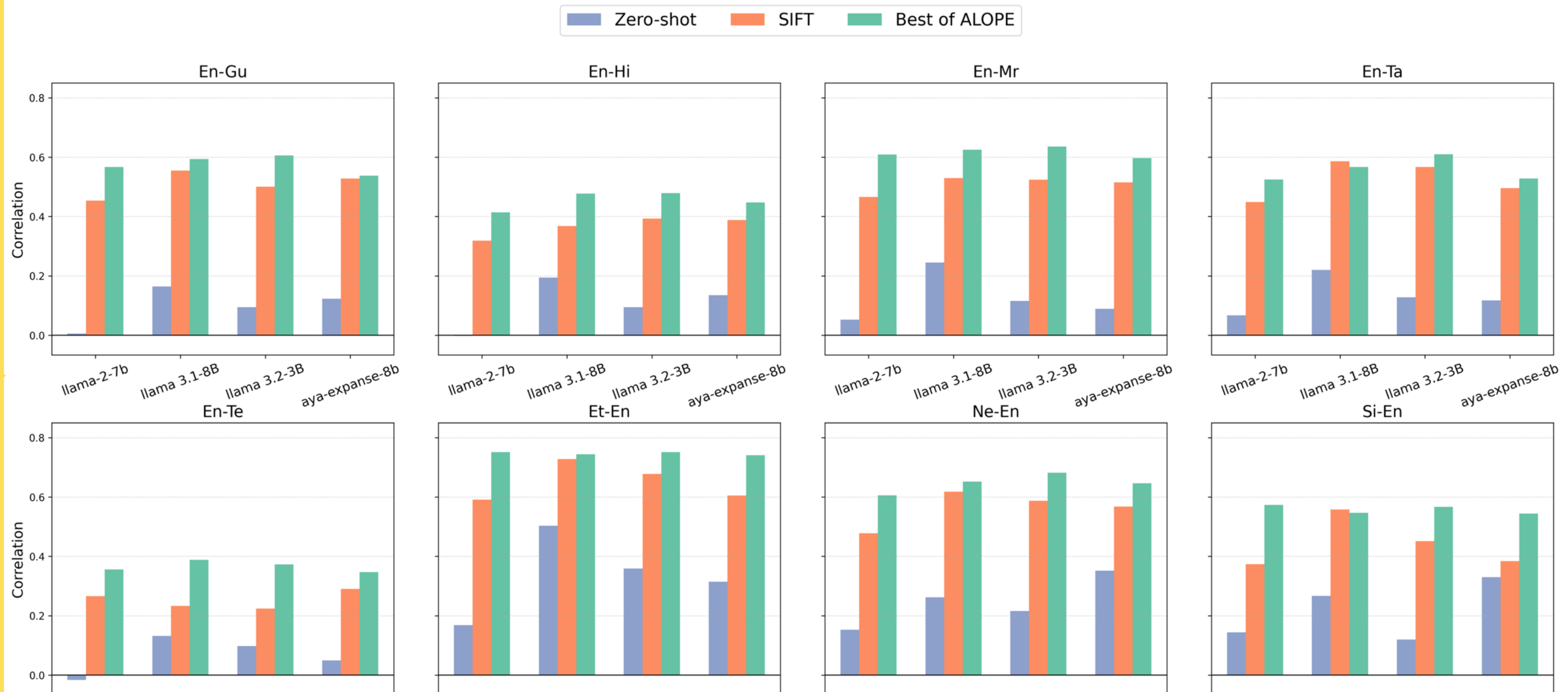
	Model	En-Gu	En-Hi	En-Mr	En-Ta	En-Te	Et-En	Ne-En	Si-En	*Avg.*
TL (-1)	**llama-2-7b**	0.563	0.414	0.609	0.525	0.356	0.742	0.596	0.565	*0.546*
	llama 3.1-8B	0.594	0.469	0.620	0.567	0.363	0.734	0.647	0.547	*0.567*
	llama 3.2-3B	0.604	0.477	**0.636**	0.580	0.348	0.735	0.674	0.543	*0.575*
	aya-expanse-8b	0.068	0.178	0.219	-0.006	0.275	0.115	0.012	0.077	*0.117*
	Avg	*0.457*	*0.385*	*0.521*	*0.416*	*0.335*	*0.581*	*0.482*	*0.433*	
TL (-7)	**llama-2-7b**	0.567	0.336	0.542	0.484	0.317	0.739	0.606	**0.573**	*0.520*
	llama 3.1-8B	0.590	0.477	0.625	0.528	**0.388**	0.744	0.638	0.544	*0.567*
	llama 3.2-3B	**0.606**	**0.479**	0.617	0.585	0.369	**0.751**	0.664	0.553	*0.578*
	aya-expanse-8b	0.538	0.447	0.597	0.528	0.347	0.741	0.646	0.544	*0.549*
	Avg	*0.575*	*0.435*	*0.595*	*0.531*	*0.355*	*0.744*	*0.639*	*0.554*	
TL (-11)	**llama-2-7b**	0.360	0.301	0.361	0.254	0.293	0.405	0.164	0.049	*0.273*
	llama 3.1-8B	0.514	0.412	0.609	0.438	0.304	0.148	0.554	0.493	*0.434*
	llama 3.2-3B	0.594	0.476	0.605	**0.610**	0.373	0.748	0.678	0.560	*0.581*
	aya-expanse-8b	0.490	0.411	0.572	0.445	0.336	0.569	0.453	0.439	*0.464*
	Avg	*0.489*	*0.400*	*0.537*	*0.437*	*0.327*	*0.467*	*0.462*	*0.385*	
TL (-16)	**llama-2-7b**	0.540	0.381	0.585	0.482	0.308	0.751	0.580	0.569	*0.524*
	llama 3.1-8B	0.558	0.453	0.602	0.523	0.350	0.737	0.652	0.513	*0.548*
	llama 3.2-3B	0.557	0.459	0.597	0.547	0.338	0.745	**0.682**	0.567	*0.561*
	aya-expanse-8b	0.467	0.390	0.557	0.481	0.314	0.727	0.576	0.540	*0.506*
	Avg	*0.530*	*0.421*	*0.585*	*0.508*	*0.327*	*0.740*	*0.622*	*0.547*	
TL (-20)	**llama-2-7b**	0.470	0.405	0.544	0.460	0.338	0.684	0.508	0.534	*0.493*
	llama 3.1-8B	0.484	0.394	0.553	0.321	0.172	0.649	0.524	0.494	*0.449*
	llama 3.2-3B	0.430	0.408	0.579	0.303	0.286	0.601	0.488	0.464	*0.445*
	aya-expanse-8b	0.437	0.300	0.488	0.263	0.287	0.483	0.438	0.395	*0.386*
	Avg	*0.455*	*0.377*	*0.541*	*0.337*	*0.271*	*0.604*	*0.490*	*0.472*	
TL (-24)	**llama-2-7b**	0.500	0.421	0.538	0.379	0.239	0.630	0.507	0.472	*0.461*
	llama 3.1-8B	0.421	0.378	0.552	0.330	0.290	0.515	0.530	0.464	*0.435*
	llama 3.2-3B	0.443	0.376	0.507	0.367	0.299	0.559	0.528	0.487	*0.446*
	aya-expanse-8b	0.375	0.319	0.440	0.337	0.220	0.393	0.407	0.345	*0.354*
	Avg	0.435	0.373	0.509	0.353	0.262	0.524	0.493	0.442	

Approaches with ALOPE vs SFIT



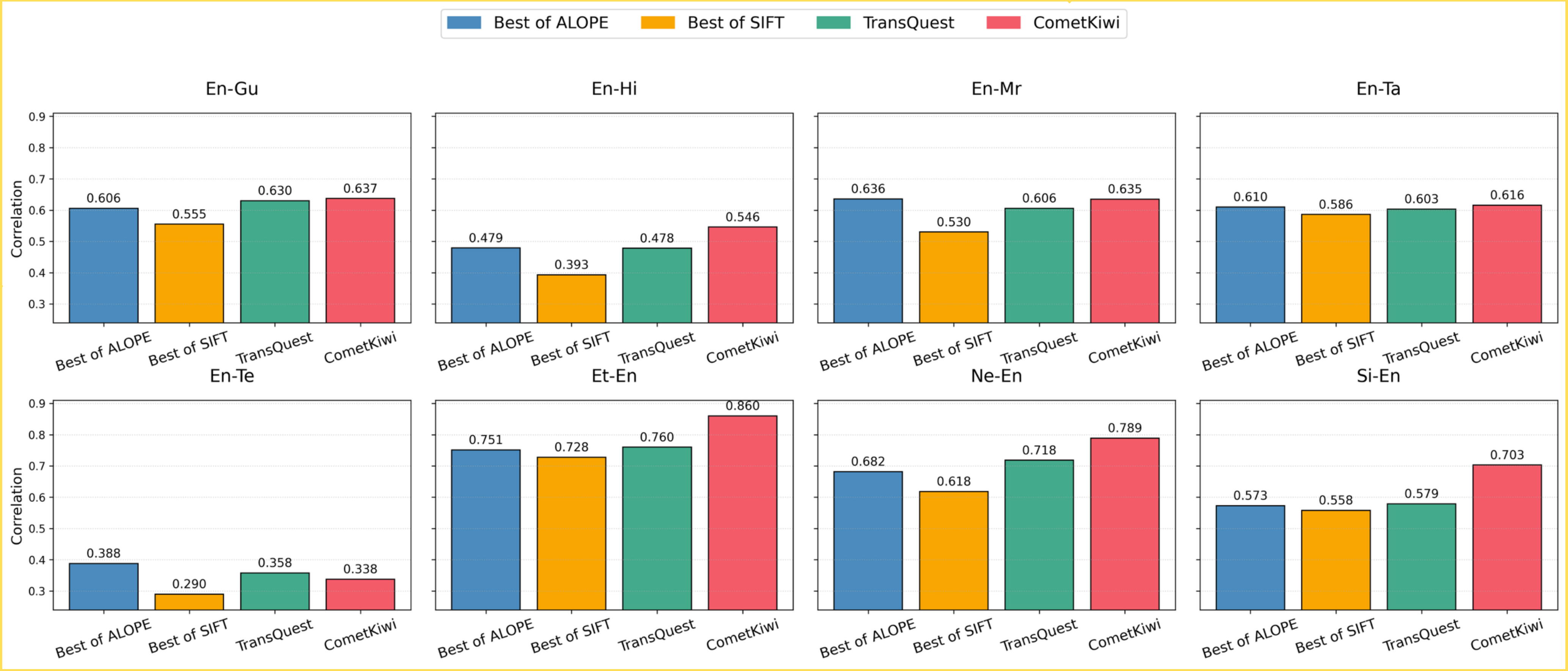
ALOPE *vs.* Zero-shot *vs.* SIFT

	Model	En-Gu	En-Hi	En-Mr	En-Ta	En-Te	Et-En	Ne-En	Si-En
Zero-shot	llama-2-7b	0.006	-0.002	0.053	0.067	-0.016	0.168	0.153	0.144
	llama 3.1-8B	**0.164**	**0.194**	**0.245**	**0.220**	**0.132**	**0.503**	0.262	0.267
	llama 3.2-3B	0.095	0.095	0.116	0.128	0.098	0.359	0.216	0.120
	aya-expanse-8B	0.123	0.135	0.089	0.117	0.049	0.315	**0.352**	**0.330**
SIFT	llama-2-7b	0.454	0.319	0.466	0.449	0.266	0.591	0.478	0.374
	llama 3.1-8B	**0.555**	0.368	**0.530**	**0.586**	0.233	**0.728**	**0.618**	**0.558**
	llama 3.2-3B	0.501	**0.393**	0.524	0.567	0.224	0.678	0.587	0.451
	aya-expanse-8B	0.528	0.388	0.515	0.496	**0.290**	0.605	0.568	0.384
Best of ALOPE	llama-2-7b	0.567	0.414	0.609	0.525	0.356	0.751	0.606	**0.573**
	llama 3.1-8B	0.594	0.477	0.625	0.567	**0.388**	0.744	0.652	0.547
	llama 3.2-3B	**0.606**	**0.479**	**0.636**	**0.610**	0.373	**0.751**	**0.682**	0.567
	aya-expanse-8B	0.538	0.447	0.597	0.528	0.347	0.741	0.646	0.544

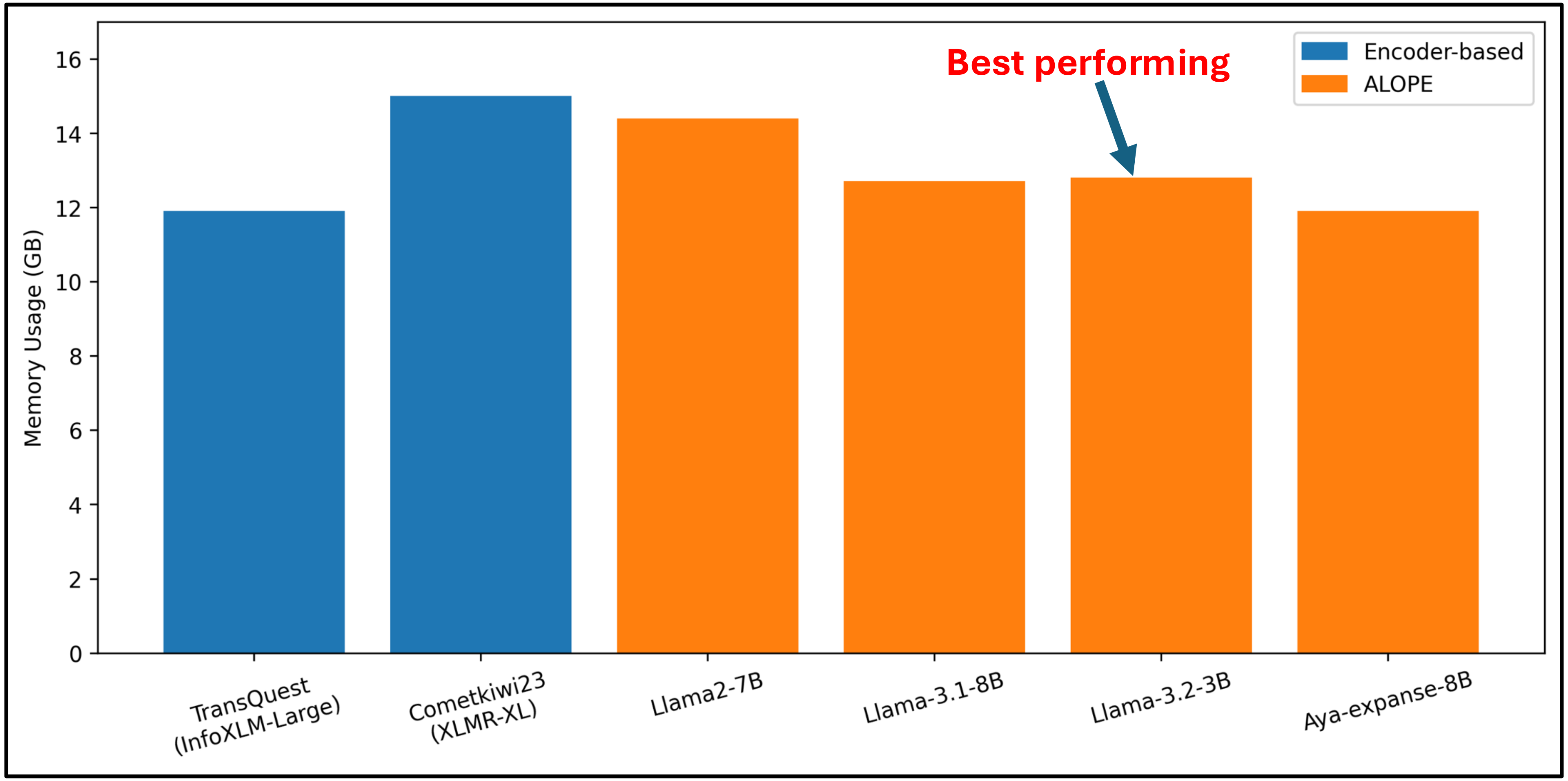


ALOPE vs SIFT vs SOTA Approaches

Approach	En-Gu	En-Hi	En-Mr	En-Ta	En-Te	Et-En	Ne-En	Si-En
Best ALOPE	0.606	0.479	**0.636**	0.610	**0.388**	0.751	0.682	0.573
Best from SIFT	0.555	0.393	0.530	0.586	0.290	0.728	0.618	0.558
TransQuest	0.630	0.478	0.606	0.603	0.358	0.760	0.718	0.579
CometKiwi	**0.637**	**0.546**	0.635	**0.616**	0.338	**0.860**	**0.789**	**0.703**



Computational Efficiency



Conclusion

- **Novel framework**
  - Regression heads + LoRA on LLMs
- **Performance**
  - Outperforms SIFT; matches / exceeds the performance of SOTA approaches QE
- **Findings**
  - Intermediate layers (TL-7, TL-11) => most effective for low-resource cross-lingual QE
  - Mid=> layers stabilize earlier when English is target
  - LLaMA-3.2 => best overall despite small size
- **Extension to the framework**
  - Dynamic weighting & multi-head regression improve baseline SFIT
- **Efficiency**
  - Competitive GPU memory , practical & scalable
- **Future vision**
  - Enhancing ALOPE for error reasoning and automatic post-editing
  - Look out for **ALOPE-RL**, and **ALOPE-APE** branches of this work! ☺

Thank You ☺

Questions?