VIS 2023

TransforLearn: Interactive Visual Tutorial for the Transformer Model

Lin Gao[1], Zekai Shao[1], Ziqin Luo[1], Haibo Hu[2], Cagatay Turkay[3], Siming Chen[1,4]

[1]School of Data Science, Fudan University

[2]School of Big Data & Software Engineering, Chongqing University

[3]Centre for Interdisciplinary Methodologies, University of Warwick

[4]Shanghai Key Laboratory of Data Science

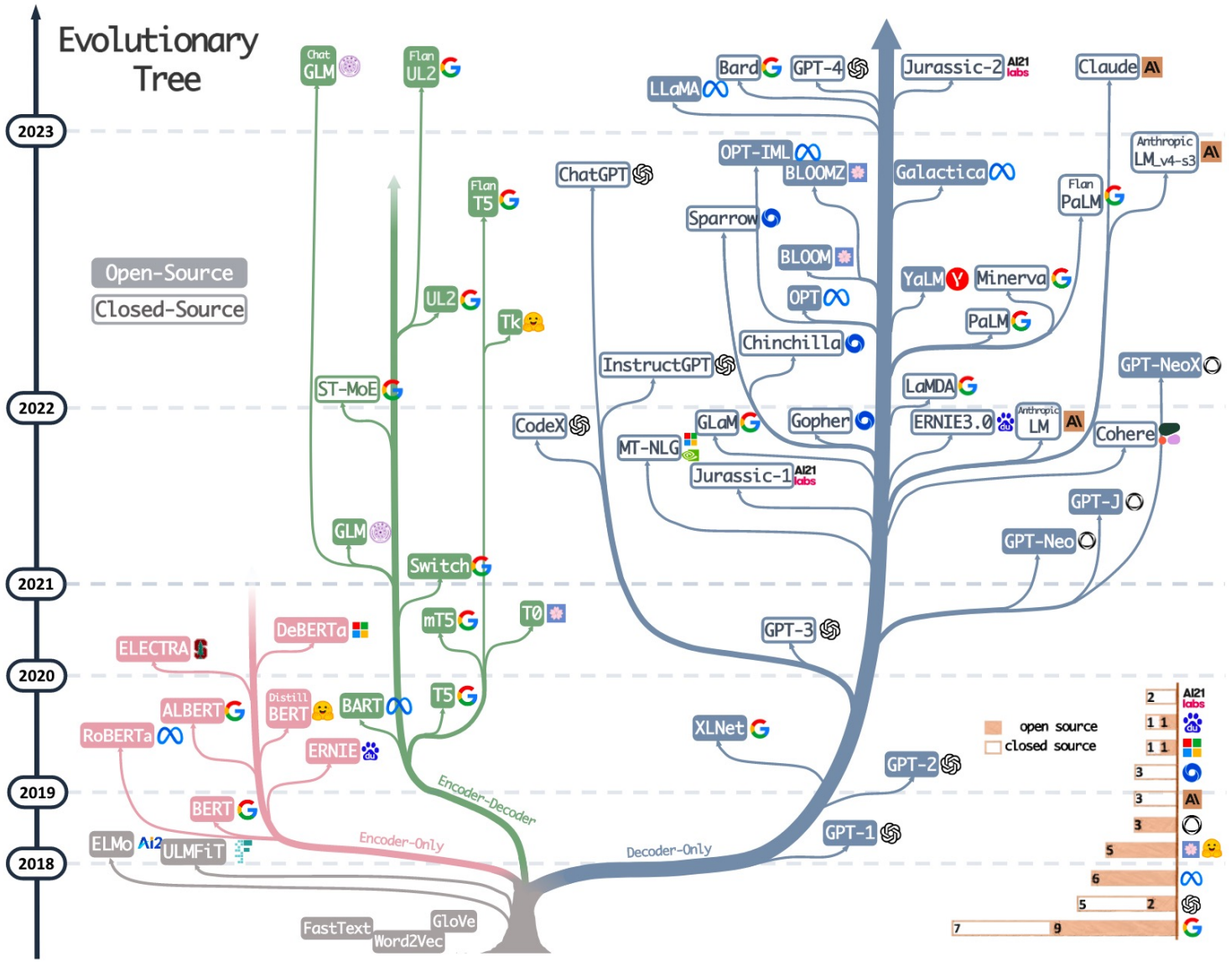





Introduction

Background

- The widespread adoption of Transformer, serving as the core framework for numerous large language models.
- The popularity of Transformer has sparked significant interest in learning its working mechanisms.



Evolutionary Tree of Transformer-based Models [1]

[1] Yang J, Jin H, Tang R, et al. Harnessing the power of llms in practice: A survey on chatgpt and beyond[J]. arXiv preprint arXiv:2304.13712, 2023.

Introduction

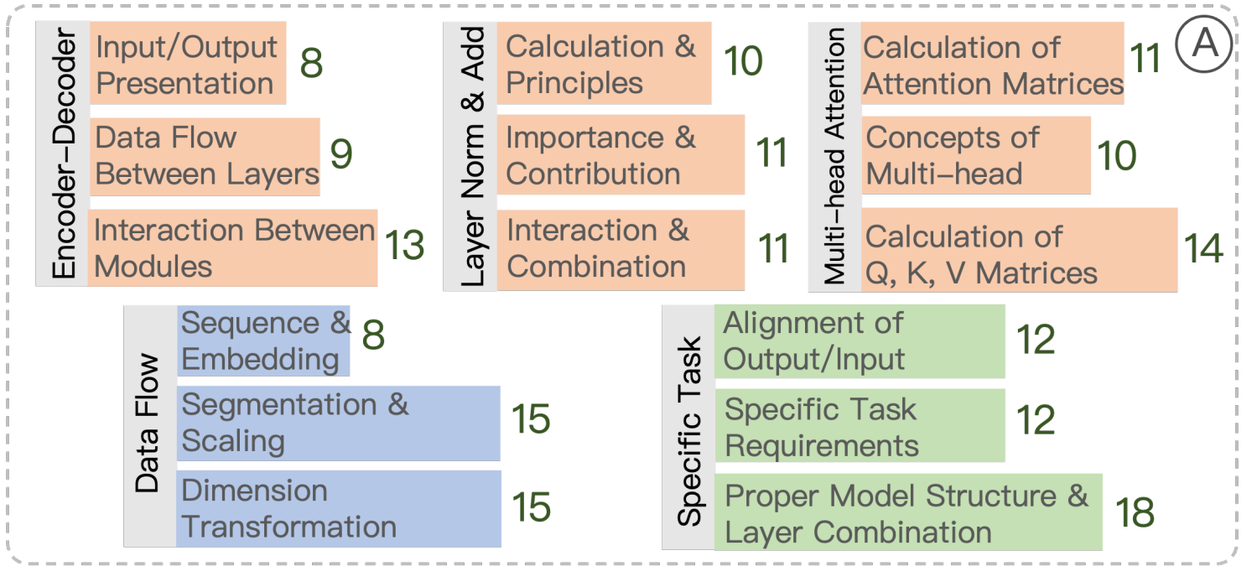
Preliminary Study

- However, beginners face difficulties in comprehending and learning Transformer due to its complex structure, data transformation and abstract downstream task.

  - Encoder/Decoder, Attention ……

  - Embedding, Dimension ……

  - Alignment, Process ……



Introduction

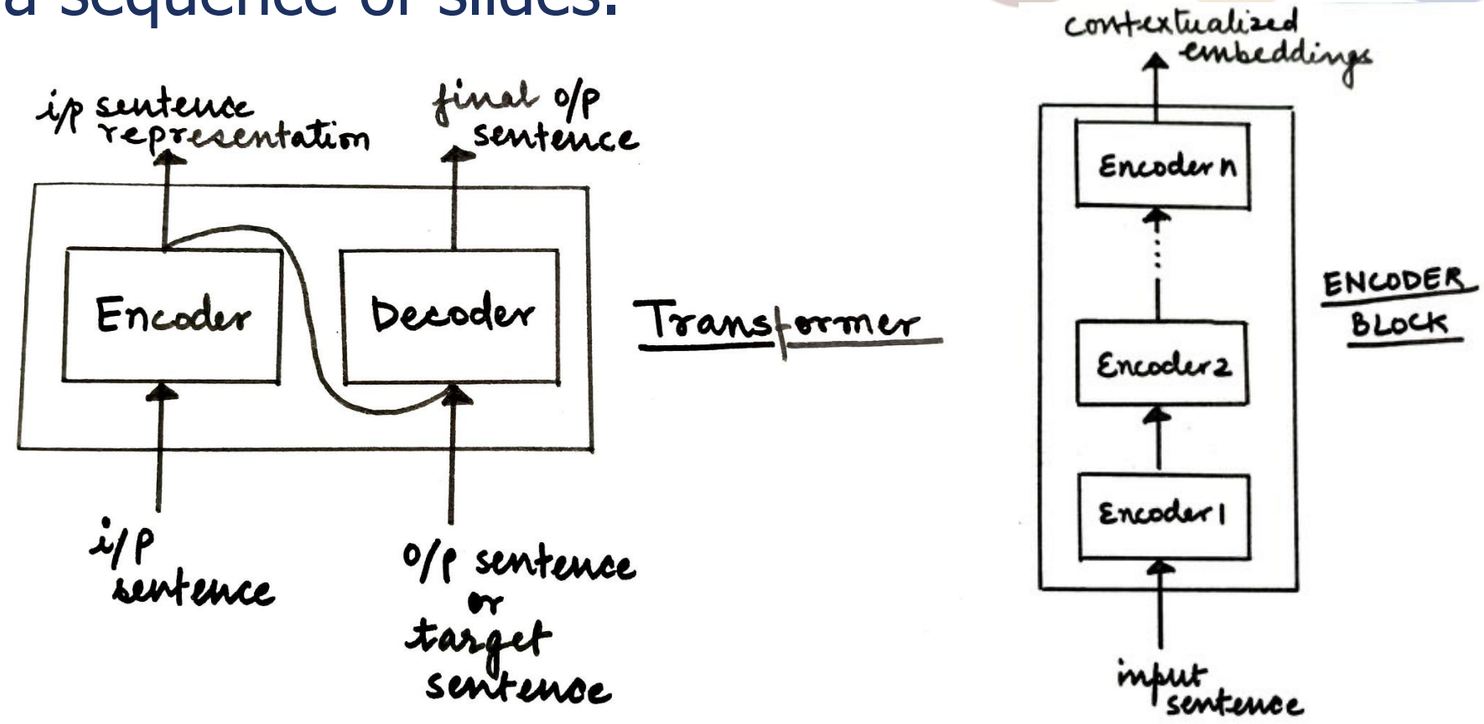
Preliminary Study

- For lecturers, they need to manually break down Transformer into multiple steps and discuss them in a sequence of slides.

  - Hands-on experience

  - Interactive resource

  - Class engagement



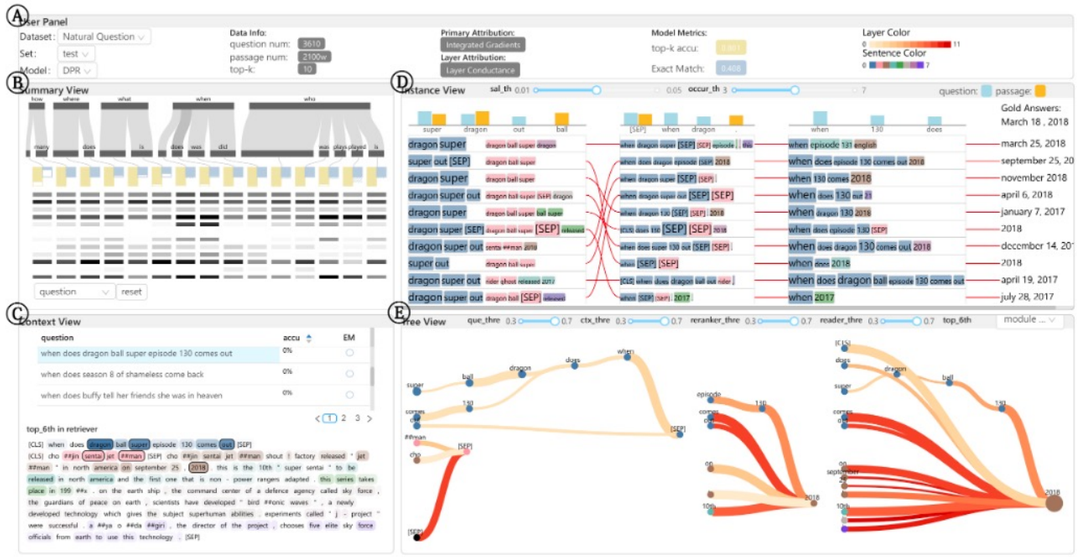
Common Way to Teach Transformers [1]

[1] https://towardsdatascience.com/transformer-models-101-getting-started-part-1-b3a77ccfa14d

Related Work

- Visualization for understanding deep learning models
  - how the models make decisions & what they learned
  - model improvement & debugging

M2lens [1] (TVCG 2021)



VEQA[2] (TVCG 2023)

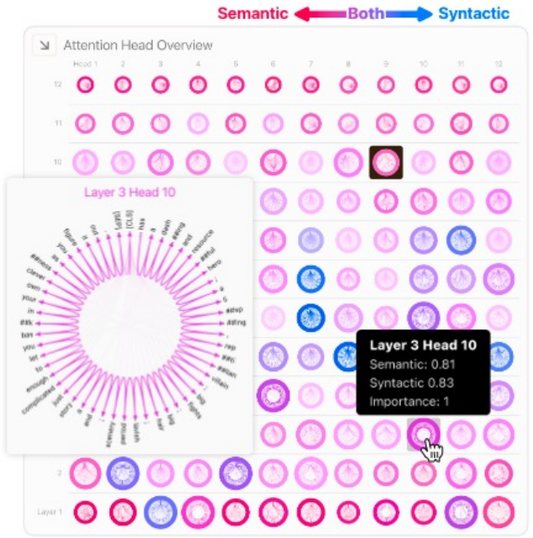
for experts, not suitable as tutorial tools

[1] Wang X, He J, Jin Z, et al. M2lens: Visualizing and explaining multimodal models for sentiment analysis[J]. IEEE Transactions on Visualization and Computer Graphics, 2021, 28(1): 802-812.

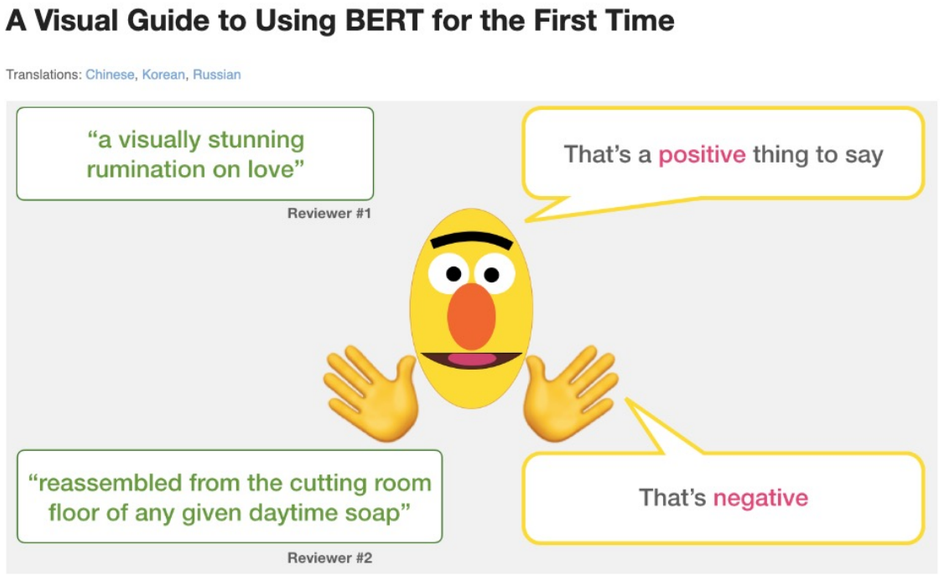
[2] Shao Z, Sun S, Zhao Y, et al. Visual Explanation for Open-domain Question Answering with BERT[J]. IEEE Transactions on Visualization and Computer Graphics, 2023.

Related Work

- Visual interpretation of Transformers
  - interpretation of embedding and attention mechanisms
  - blogs & videos for tutorial



Dodrio[1] (ACL 2021)



Jalammar's blogs [2]

lack of mathematical details

& interaction with the actual data flow or task

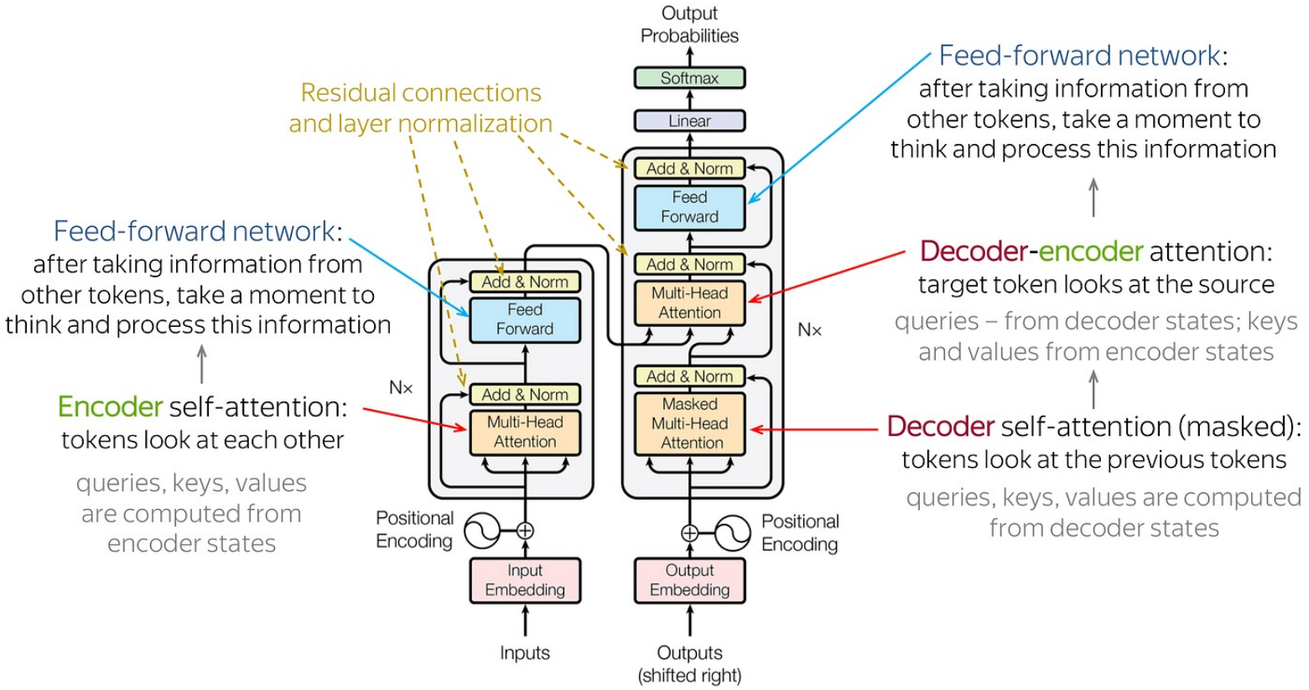
[1] Wang Z J, Turko R, Chau D H. Dodrio: Exploring transformer models with interactive visualization[J]. arXiv preprint arXiv:2103.14625, 2021.
[2] https://jalammar.github.io/a-visual-guide-to-using-bert-for-the-first-time/

Requirement Analysis

Tasks & Goals

Consequently, an interactive visual tutorial is needed to for deep learning beginners and non-experts to comprehensively learn about Transformers.

What can TransforLearn do?

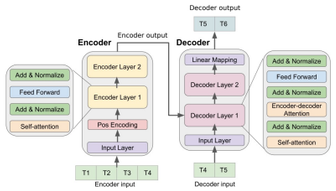


Why does Transformer has such a complex architecture[1]

[1] https://stats.stackexchange.com/questions/512242/why-does-transformer-has-such-a-complex-architecture

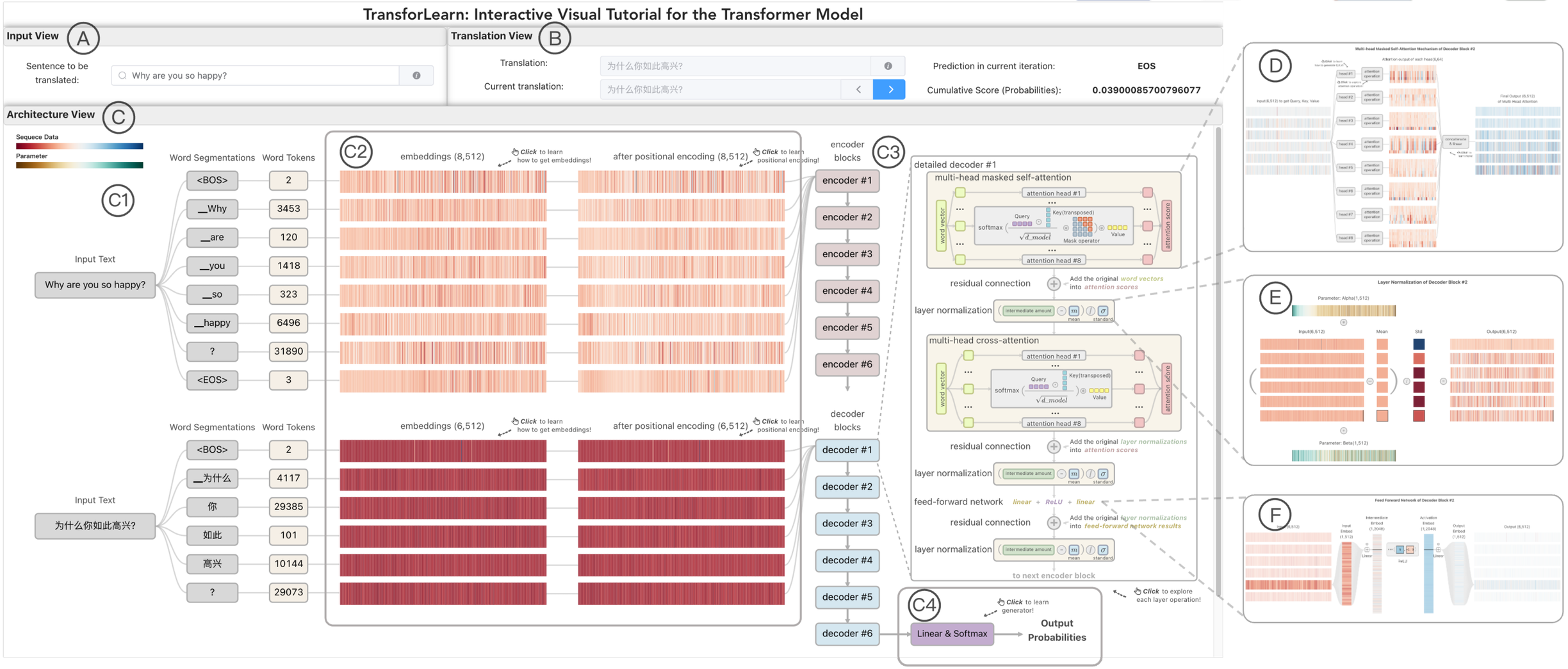
Requirement Analysis

Tasks & Requirements

Task-1	Task-2	Task-3	Task-4
			
complex structure & layer operations	data flow & transformation	practical use in downstream tasks	guidance & feedback
Requirement-1	Requirement-2	Requirement-3	Requirement-4
A visual summary of the model architecture and data flow.	An interactive interface for layer operations and mathematical formulas.	Exploration mode between module levels based on downstream tasks.	Self-directed and immersive learning experiences.

Visual Design

Overview



Visual Design

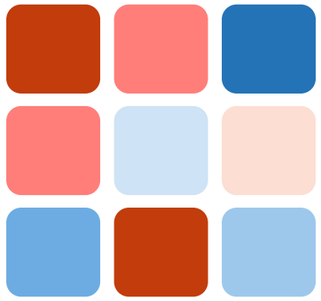
Overview

Architecture Overview

$$\begin{bmatrix} 11 & 20 & 109 \\ 21 & 54 & 37 \\ 74 & 11 & 60 \end{bmatrix}$$

Sequence Data
Parameter Data

Module Detailed Views



Breaking text into individual word segmentations.

Word Segmentations

Mapping words to dense vector representations.

embeddings (4,512)

Index in word token dictionary.

Word Tokens

Add positional information to original embeddings.

after positional encoding (4,512)

Add the original *layer normalizations* into *attention scores*

Add the original *layer normalizations* into *feed-forward network results*

Click to explore each layer operation!

Click to learn how to generate Q,K,V!

Visual Design

Overview

Attention Mechanism

⇩

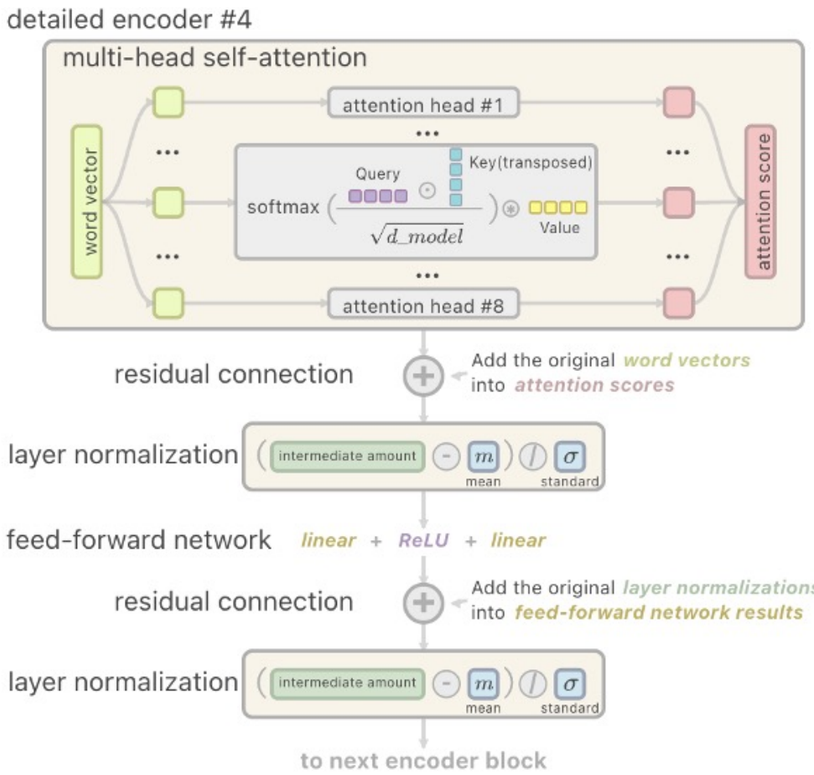
Add + Layer Normalization

⇩

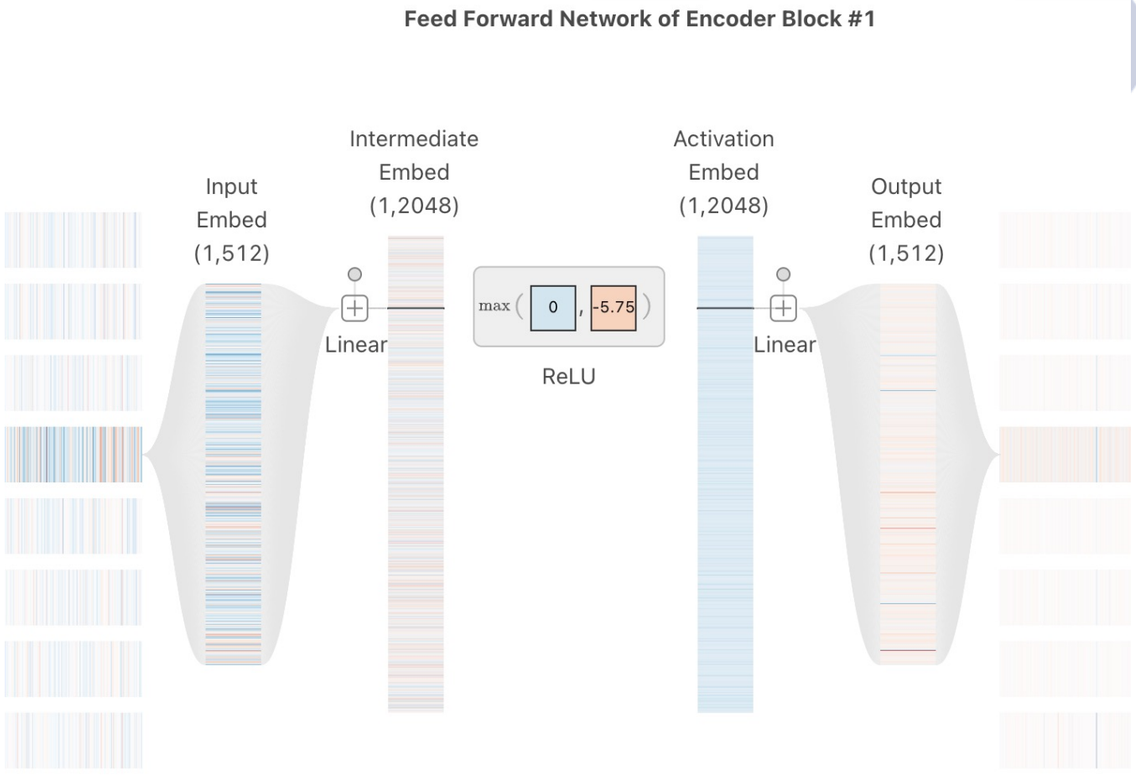
Feed-forward Network

⇩

Add + Layer Normalization



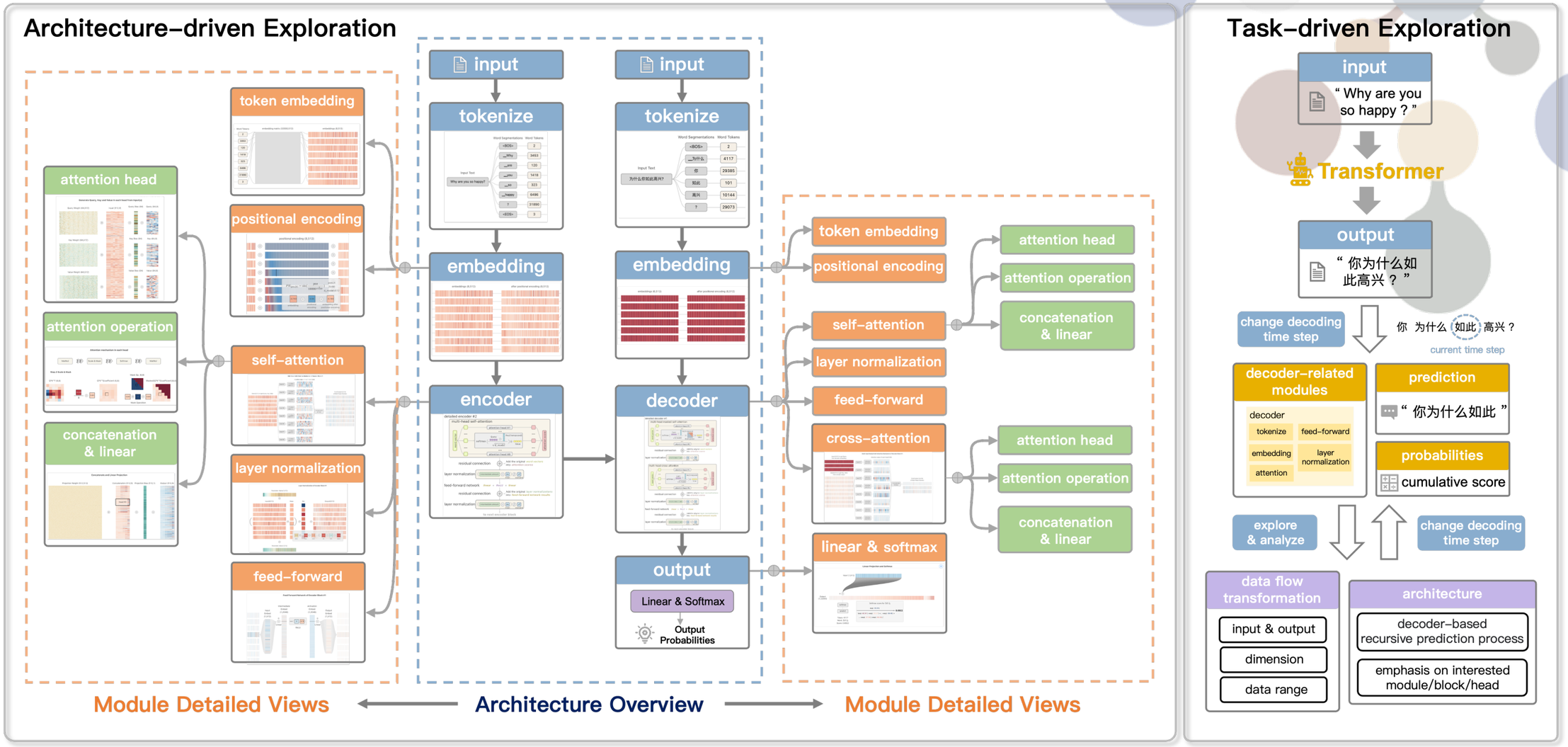
Encoder Block



Feed-forward Network

Visual Design

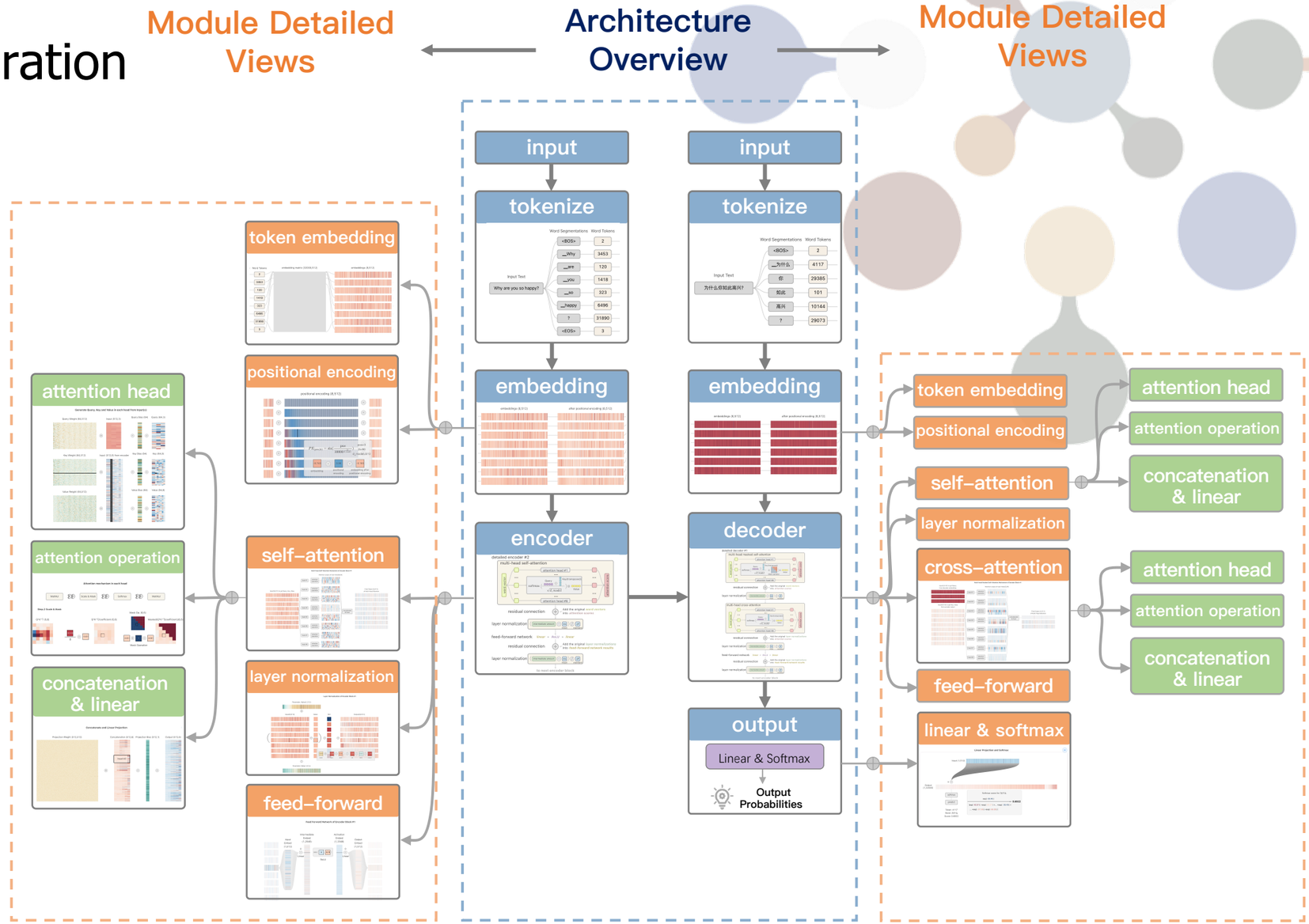
Overview



Visual Design

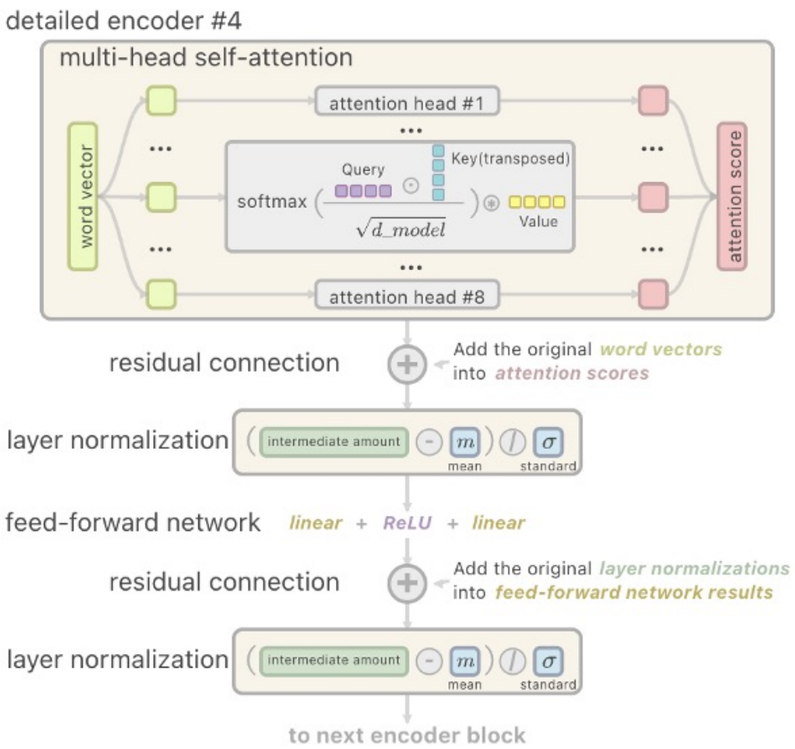
Architecture-driven Exploration

- Model overview and details
  - Hierarchical structure
  - Layer operations
  - Mathematical formulas

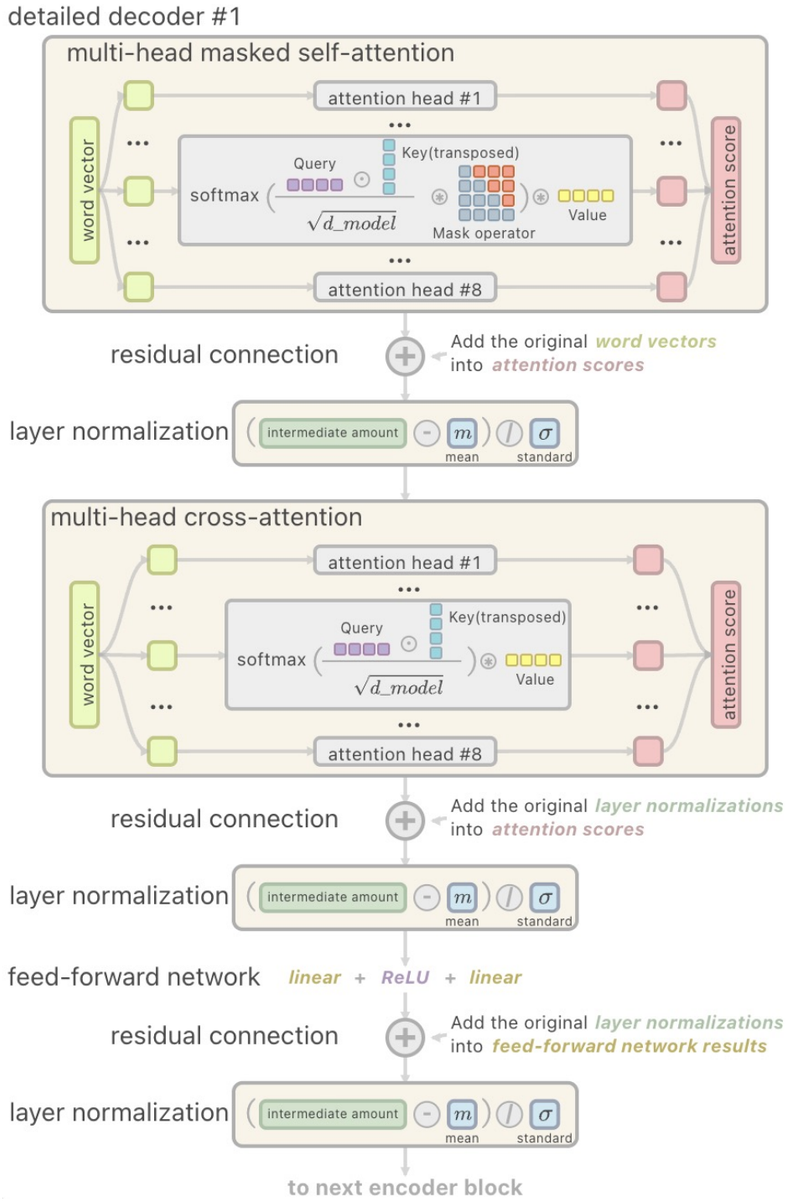


Visual Design

Architecture-driven Exploration



Encoder Block



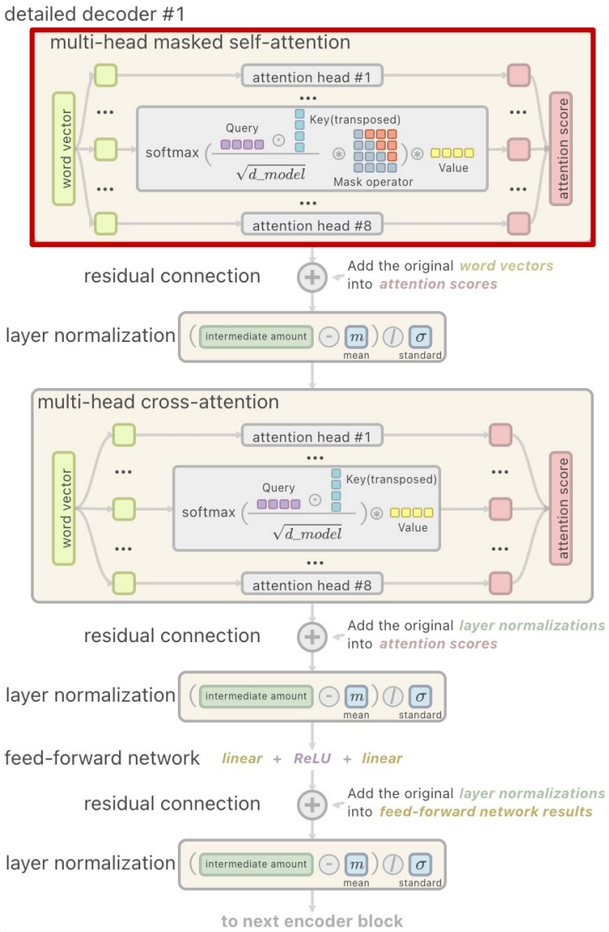
Decoder Block

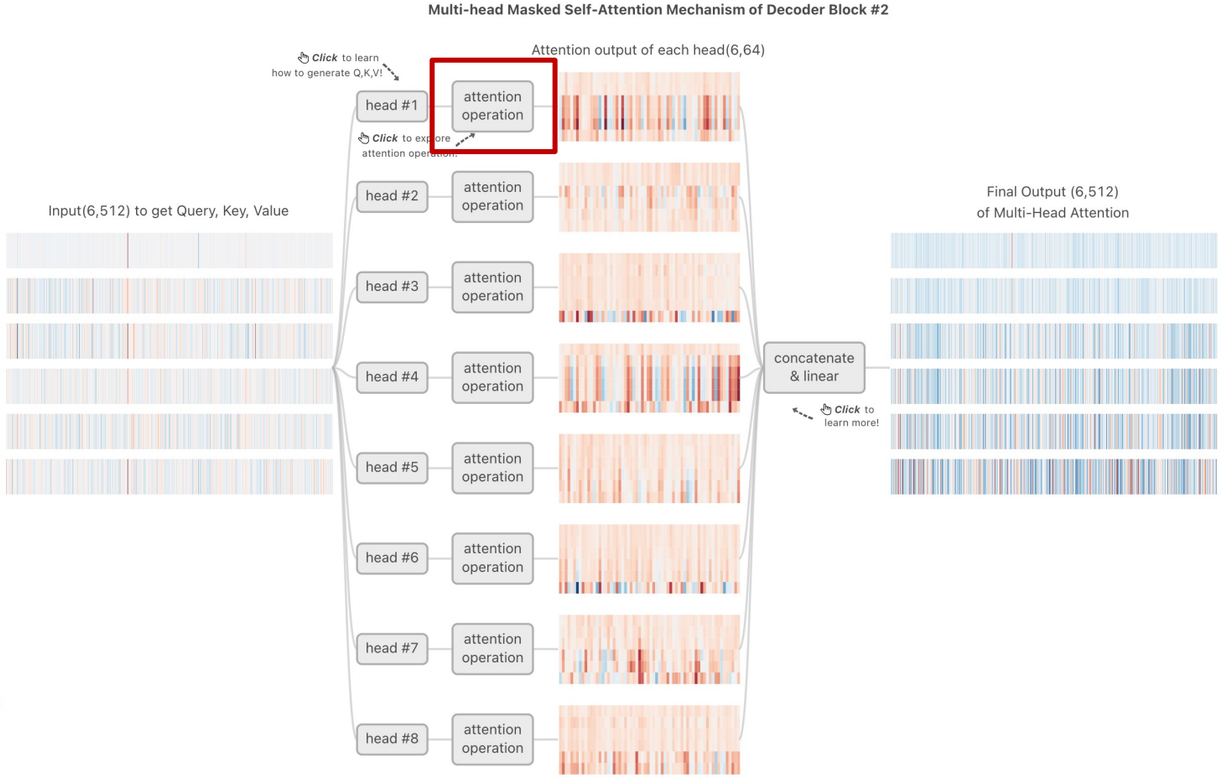
Visual Design

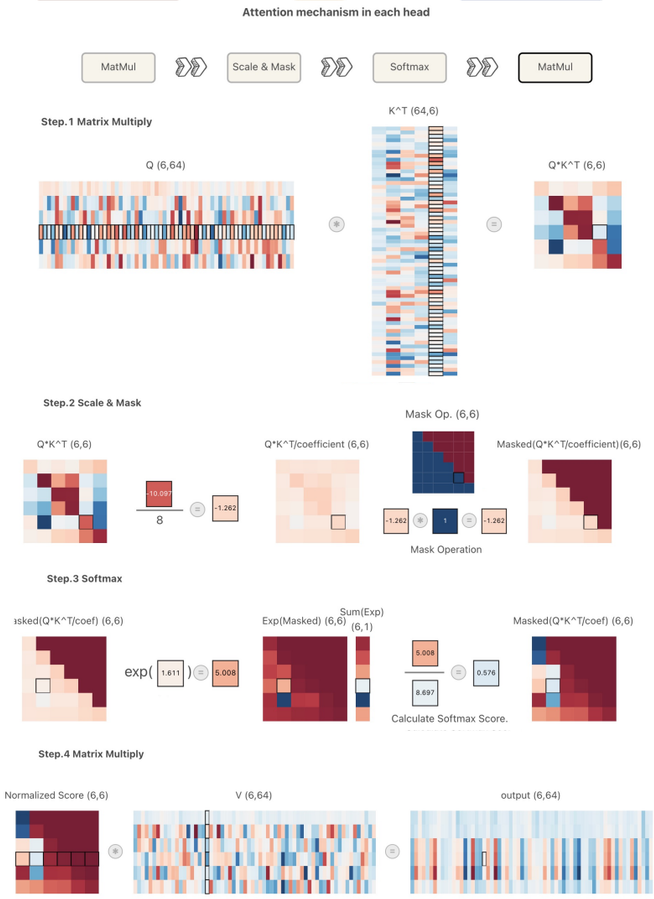
Architecture-driven Exploration

Click on **Attention**

Click on **Attention Operation**





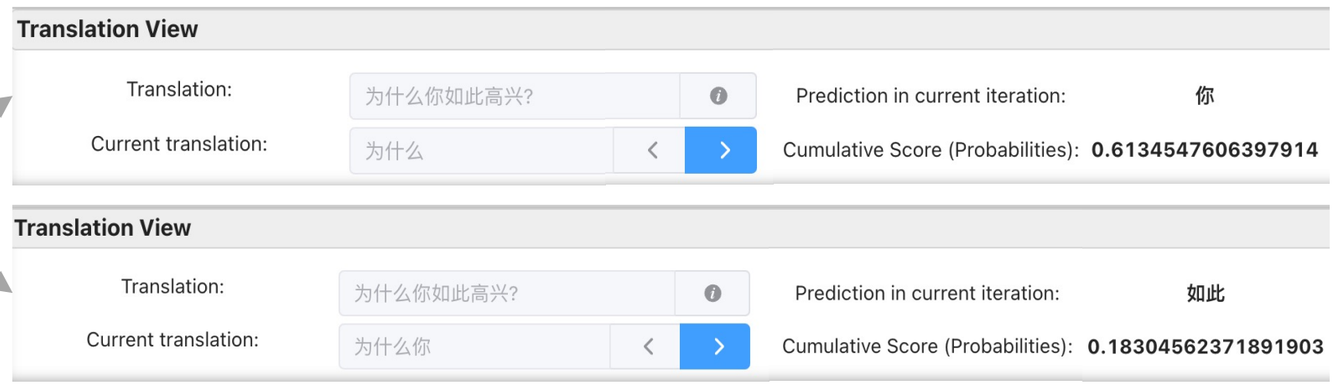


Visual Design

Task-driven Exploration

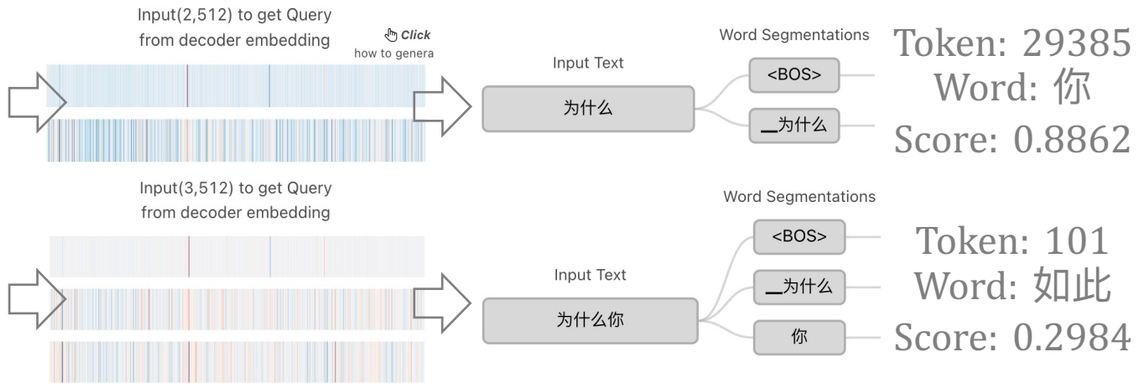
- Explore data flow changes
  - Input and output, data dimension, data range
- Analyze structural features
  - Decoding time step -> translation progress
  - Focus on a specific module or head

Click to change
the decoder time step



data flow transformation

architecture



Visual Design

Usage Scenario

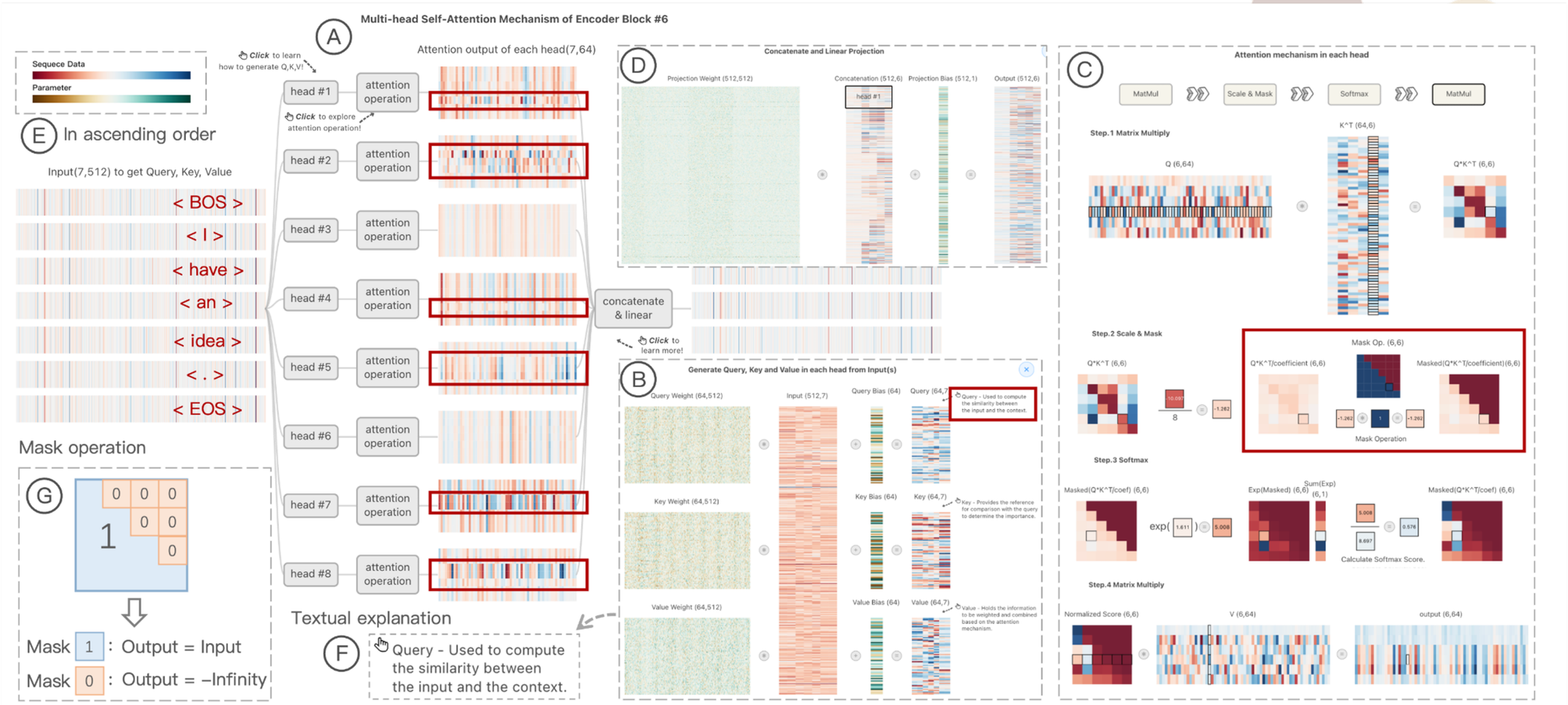
- Self-study guidance for a beginner
  - utilize Transformer to extract features from sequence data
  - the concept and generation process of the Q, K, and V matrices
  - the use of decoders for prediction

- Teaching aid for lectures
  - better summarize and present the teaching points
  - increases the practicality and vividness of the entire teaching process

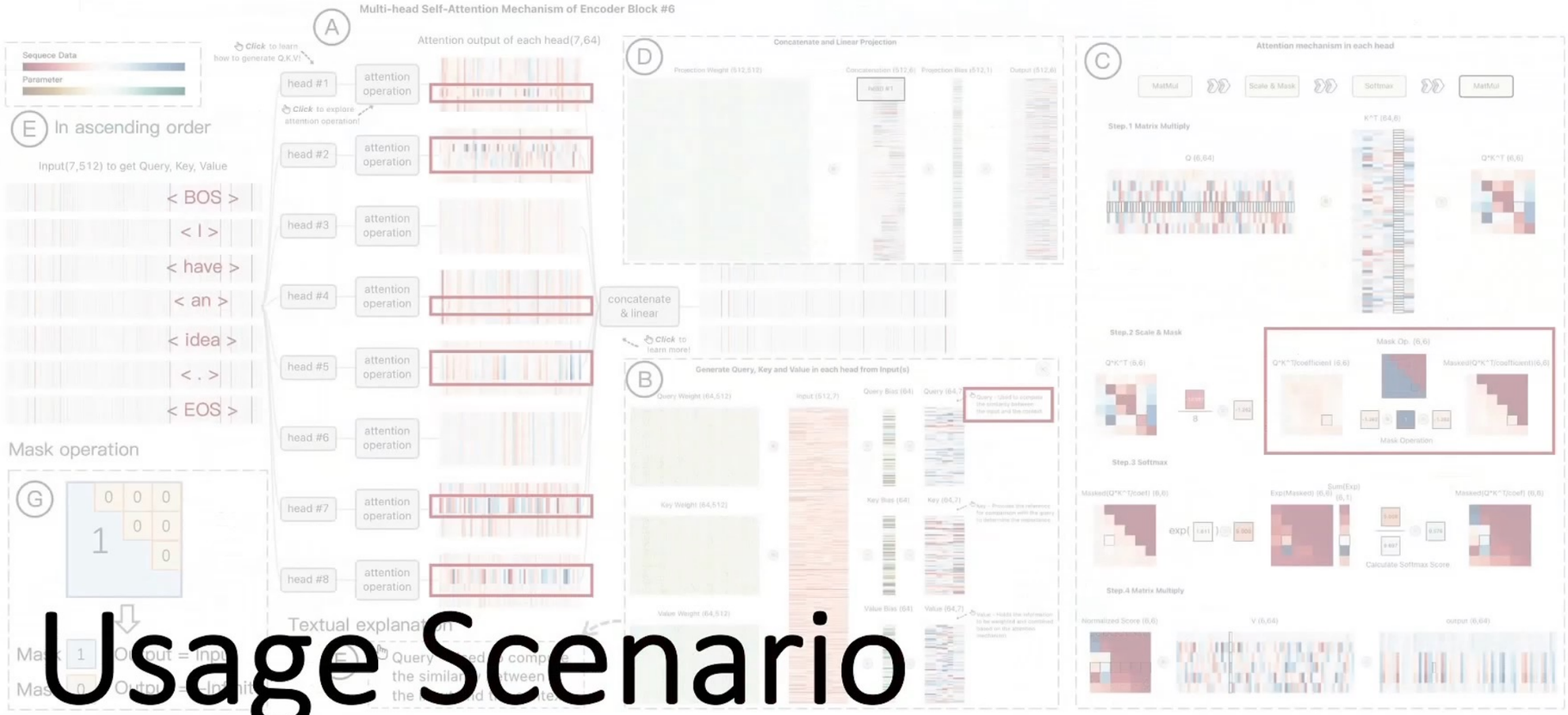
Visual Design

Usage Scenario

- Self-study guidance for a beginner



Visual Design



Usage Scenario

Self-study guidance for beginners

Evaluation

- User-controlled experiment (Objective)

User Portrait

System Introduction

System Tutorial

Observable Study

Exit Questionnaire

Requirement-1 visual summary

Requirement-2 interactive interface

Requirement-3 exploration mode

Requirement-4 self-directed & immersive

Level	Goal	Question
easy	G1	Q1: Components and data flow of feed-forward network.
easy	G3	Q2: Identify key words from attention matrix.
easy	G3	Q3: Final output in translation task and its derivation.
medium	G1	Q4: Differences between cross- and self-attention.
medium	G2	Q5: Add & LN significance and implementation.
medium	G1	Q6: Parallelism in Transformer.
hard	G2	Q7: Reasons for scaling before softmax.
hard	G2	Q8: Process of calculating PE & variation with position.

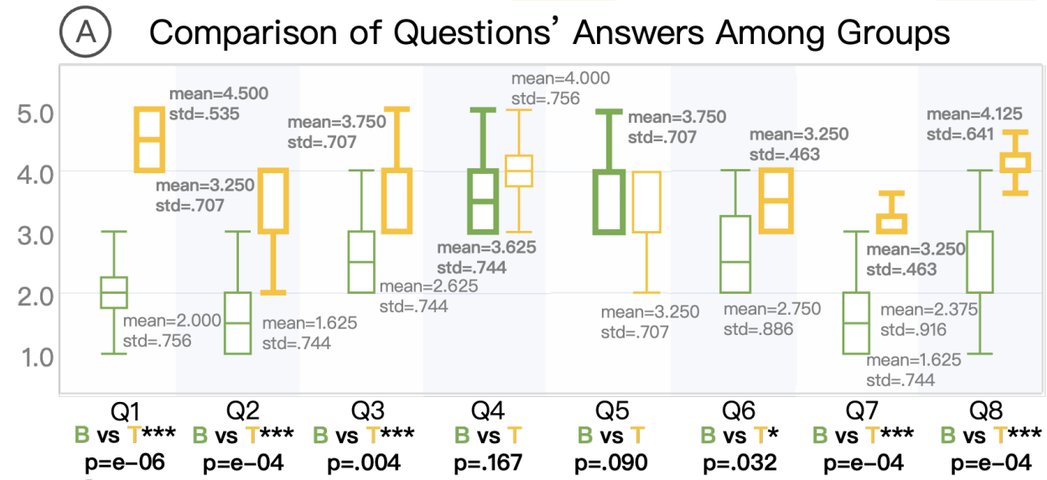
Evaluation

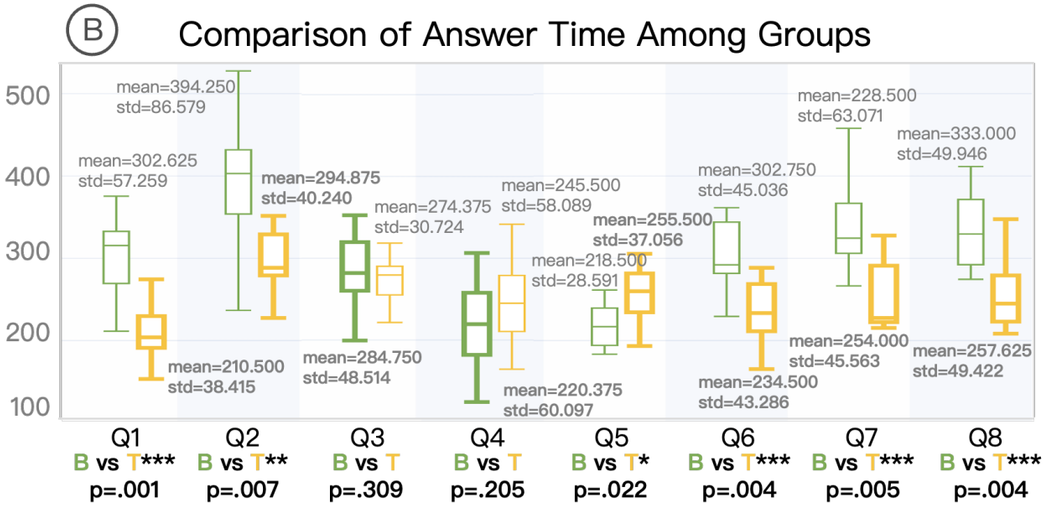
- User-controlled experiment (Objective)

  - improve users' understanding of structures and tasks

  - bring more activity, autonomy and divergent thinking

  - enhancing users' efficiency in learning through a broader coverage and enhanced interaction





C Comparison of Learning Efficiency Index Among Groups

$E_{GroupX,i}$	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$	$i=6$	$i=7$	$i=8$	Mean	Std
$X=$ B	0.737	1.005	0.680	0.839	0.981	0.824	0.965	0.851	0.851	0.121
$X=$ T	1.421	1.702	1.631	1.402	1.263	1.385	1.381	1.542	**1.466**	0.146

Evaluation & Discussion

- User interviews (Subjective)
  - Implication
    - Usability and effectiveness.
    - Validating the knowledge for experts.

  - Limitation
    - Different appropriate learning resources for different needs.
    - Need for more instructions, animations, and comparisons.

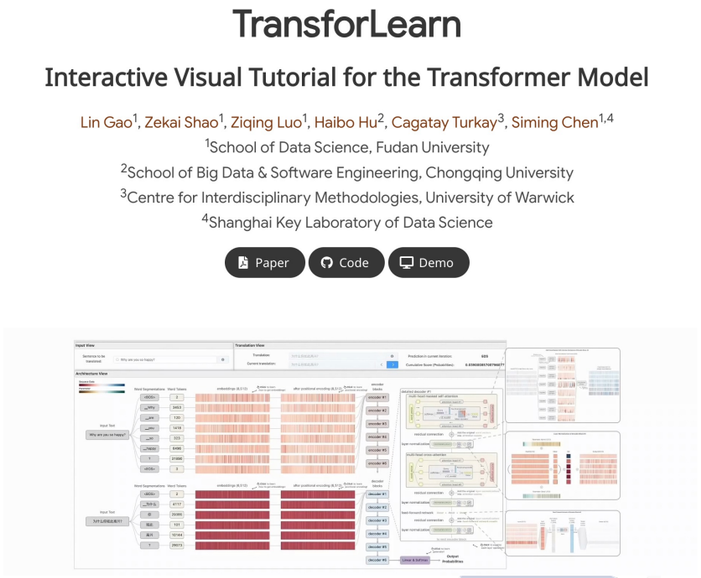
VIS 2023

Thanks for your listening ~

TransforLearn: https://trans-for-learn.github.io/

Welcome to our homepage: http://fduvis.net/

Email: leenagao0430@gmail.com



https://trans-for-learn.github.io/