



National Technical  
University of Ukraine  
"Igor Sikorsky  
Kyiv Polytechnic Institute"



Institute of  
Physics and  
Technology

# Intellectual Data Analysis

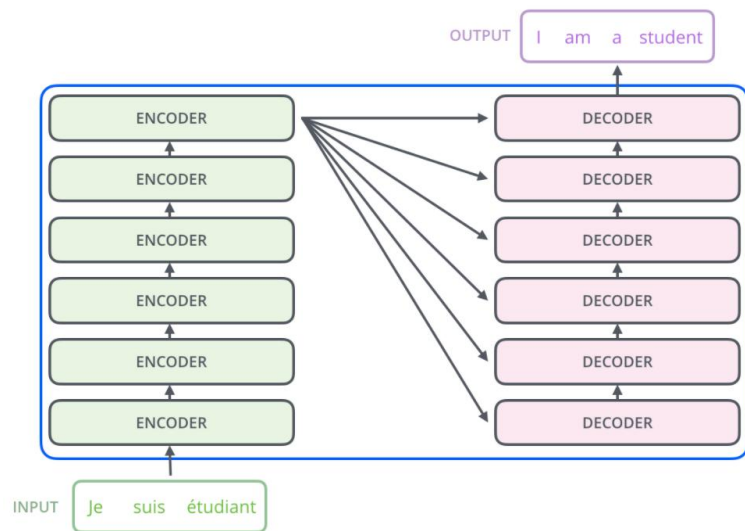
## Practice 7: Transformers

---

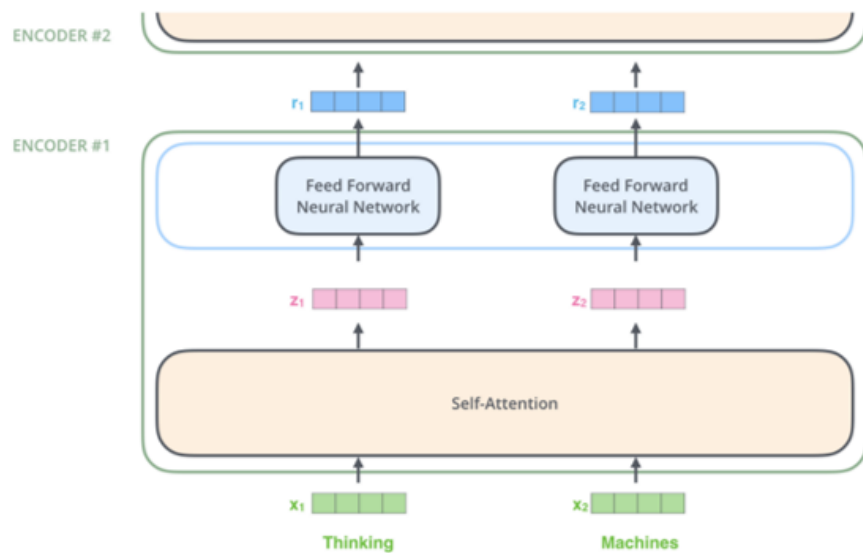
Dr. Nataliya K. Sakhnenko

*Please review Lecture 11 before this practical*

# Transformer



6 Dec 2017



## Attention Is All You Need

Ashish Vaswani\*  
Google Brain  
avaswani@google.com

Noam Shazeer\*  
Google Brain  
noam@google.com

Niki Parmar\*  
Google Research  
nikip@google.com

Jakob Uszkoreit\*  
Google Research  
usz@google.com

Llion Jones\*  
Google Research  
llion@google.com

Aidan N. Gomez\*<sup>†</sup>  
University of Toronto  
aidan@cs.toronto.edu

Lukasz Kaiser\*  
Google Brain  
lukaszkaizer@google.com

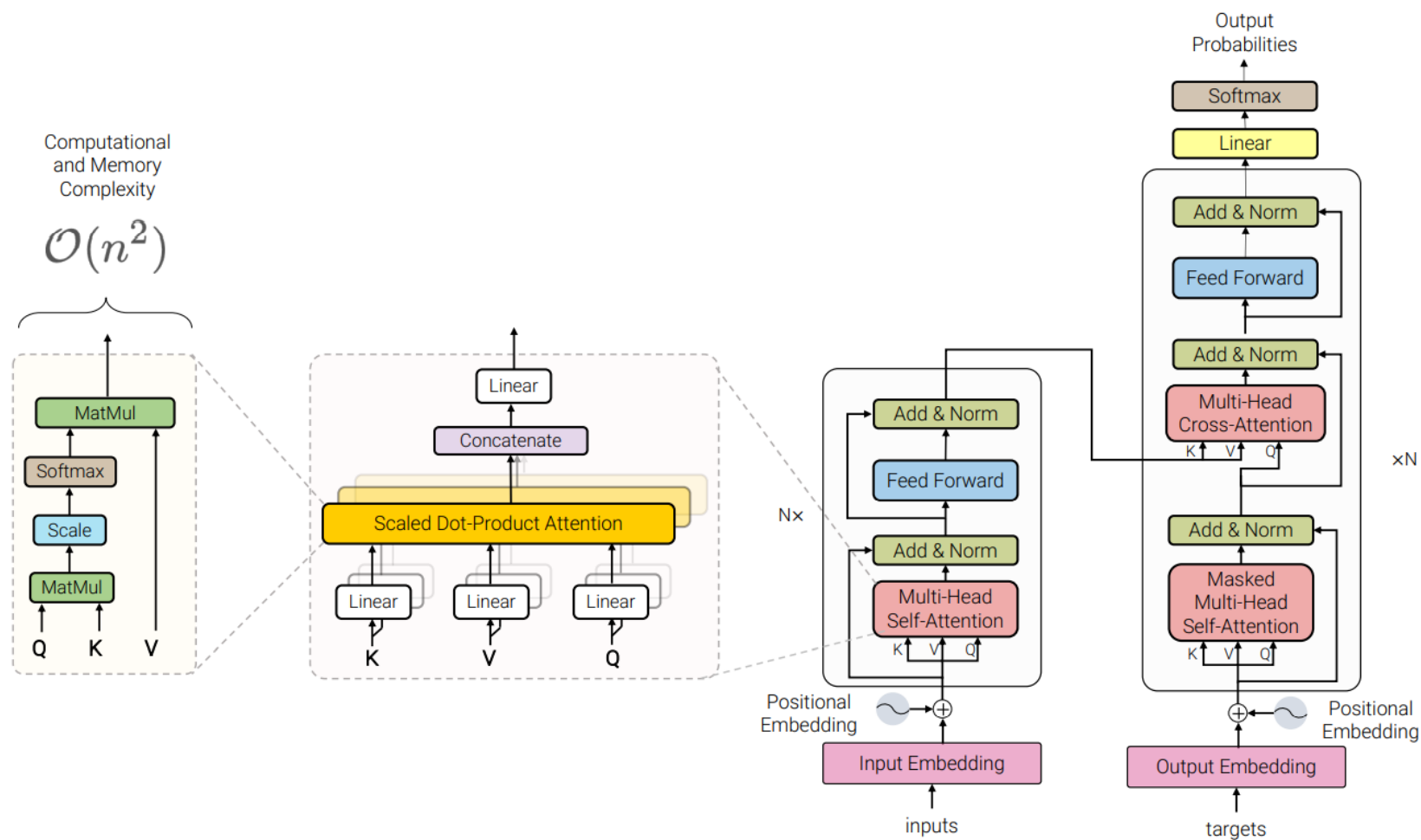
Illia Polosukhin\*<sup>‡</sup>  
illia.polosukhin@gmail.com

$$\text{softmax}\left(\frac{Q \times K^T}{\sqrt{d_k}}\right) V$$

$$= \text{Attention}(Q, K, V)$$

The diagram shows the matrix multiplication of Q (purple 3x3 matrix) and K<sup>T</sup> (orange 3x3 matrix), followed by a softmax operation. The result is then multiplied by V (blue 3x3 matrix). The final result is labeled as Z (pink 3x3 matrix) and is also labeled as Attention(Q, K, V).

# Transformer



Positional Encoding

$$PE_{(pos, 2i)} = \sin(pos/10000^{2i/d_{\text{model}}})$$

$$PE_{(pos, 2i+1)} = \cos(pos/10000^{2i/d_{\text{model}}})$$

# Hugging Face

## What is Hugging Face?

- A leading open-source ecosystem for modern AI and NLP.
- Provides thousands of **pretrained models** (Transformers, Diffusion, Speech, Vision).
- Offers large **public datasets** and tools for easy data loading.
- Includes powerful libraries: **Transformers, Datasets, Tokenizers, Accelerate**.
- Supports **fine-tuning**, evaluation, and deployment of AI models.
- Hosts **Hugging Face Hub** — a collaborative platform for sharing models, datasets, and demos.
- Enables easy experimentation via **pipelines**, Spaces, and inference widgets.
- A community-driven platform fostering **open, transparent, and reproducible AI**.

<https://huggingface.co/blog/proflead/hugging-face-tutorial>

<https://huggingface.co/learn/llm-course/en/>

LLM Course

Search documentation

Ctrl+K

EN

3,523

0. SETUP

1. TRANSFORMER MODELS

Introduction

Natural Language Processing and Large Language Models

Transformers, what can they do?

How do Transformers work?

How Transformers solve tasks

Transformer Architectures

Quick quiz

Inference with LLMs

Bias and limitations

Summary

Certification exam

2. USING TRANSFORMERS

3. FINE-TUNING A PRETRAINED MODEL

4. CHANGING MODELS AND TASKS

The most basic object in the 🤗 Transformers library is the `pipeline()` function. It connects a model with its necessary preprocessing and postprocessing steps, allowing us to directly input any text and get an intelligible answer:

```
from transformers import pipeline

classifier = pipeline("sentiment-analysis")
classifier("I've been waiting for a HuggingFace course my whole life.")
```

```
[{'label': 'POSITIVE', 'score': 0.9598047137260437}]
```

We can even pass several sentences!

```
classifier(
    ["I've been waiting for a HuggingFace course my whole life.", "I hate this so much!"]
)
```

```
[{'label': 'POSITIVE', 'score': 0.9598047137260437},
 {'label': 'NEGATIVE', 'score': 0.9994558095932007}]
```

By default, this pipeline selects a particular pretrained model that has been fine-tuned for sentiment analysis in English. The model is downloaded and cached when you create the `classifier` object. If you rerun the command, the cached model will be used instead and there is no need to download the model again.

[https://huggingface.co/docs/transformers/main\\_classes/pipelines](https://huggingface.co/docs/transformers/main_classes/pipelines)

**Hugging Face Tasks page:** [huggingface.co/tasks](https://huggingface.co/tasks)

<https://huggingface.co/learn/llm-course/en/chapter1/3>

**Model Hub (filter by Task):** [huggingface.co/models](https://huggingface.co/models)