Applied Text Analytics & Natural Language Processing

with Dr. Mahdi Roozbahani
& Wafa Louhichi

Transformers Introduction – Part 1

Learning Objectives

In this lesson, you will learn about transformers and why they are used

- Introduction
- Encoder block
- Decoder block
- Transformer structure

Transformers History

The transformers below are all pre-trained as *language models*.

- **Language model**: a model trained on a large amount of raw text using self-supervised learning, develops a statistical understanding of the language it has been trained on
- **Pre-trained model:** trained from scratch, fine-tuned using *transfer learning*
- **Self-supervised learning**: training a model on an unlabeled dataset

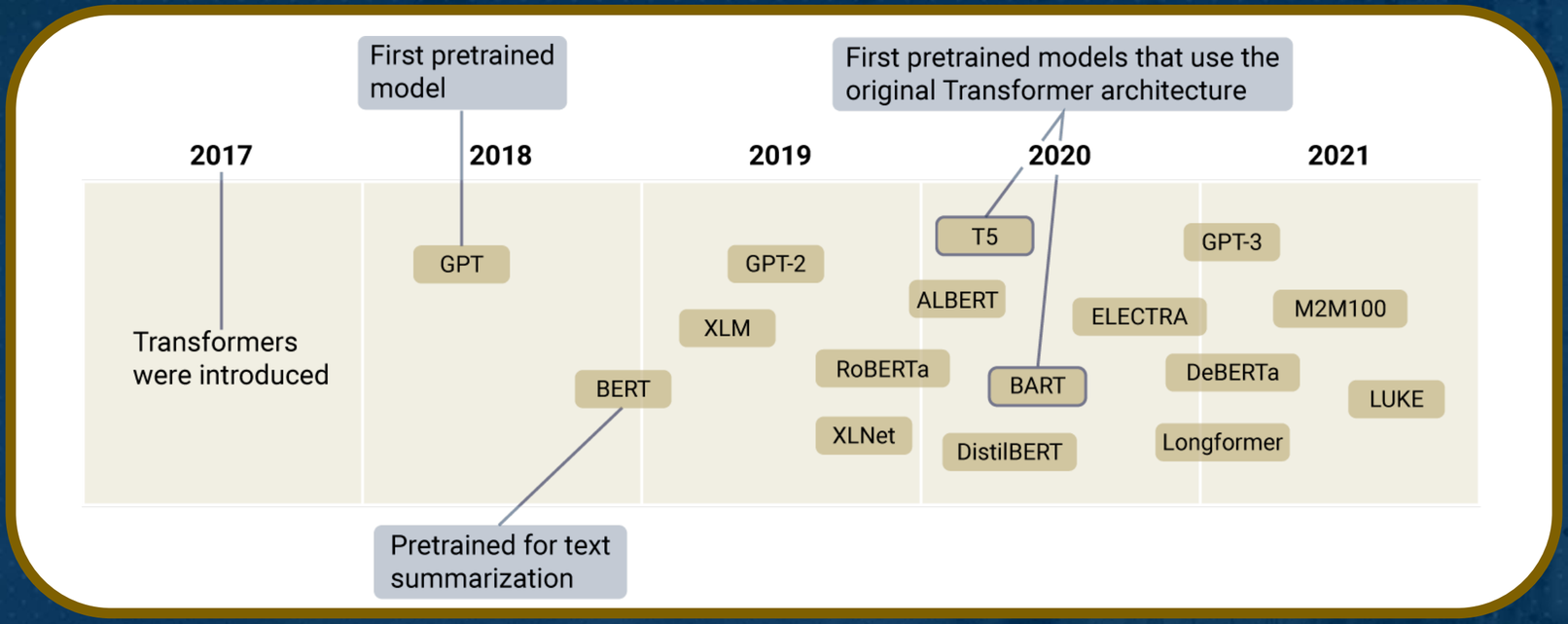


Image from Hugging Face website, edited by author https://huggingface.co/course/chapter1/1

Transformers History

The transformers below are all pre-trained as *language models*.

- **Language model**: a model trained on a large amount of raw text using self-supervised learning, develops a statistical understanding of the language it has been trained on
- **Pre-trained model:** trained from scratch, fine-tuned using *transfer learning*
- **Self-supervised learning**: training a model on an unlabeled dataset

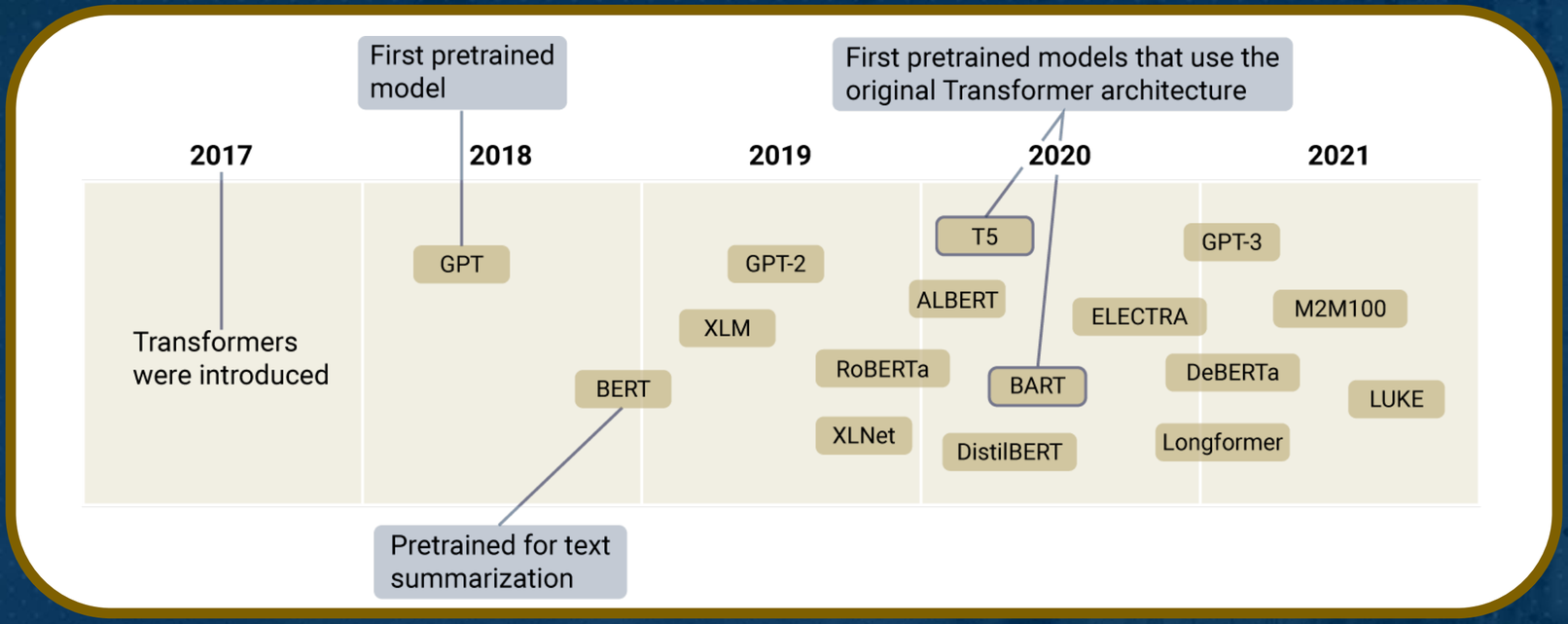


Image from Hugging Face website, edited by author https://huggingface.co/course/chapter1/1

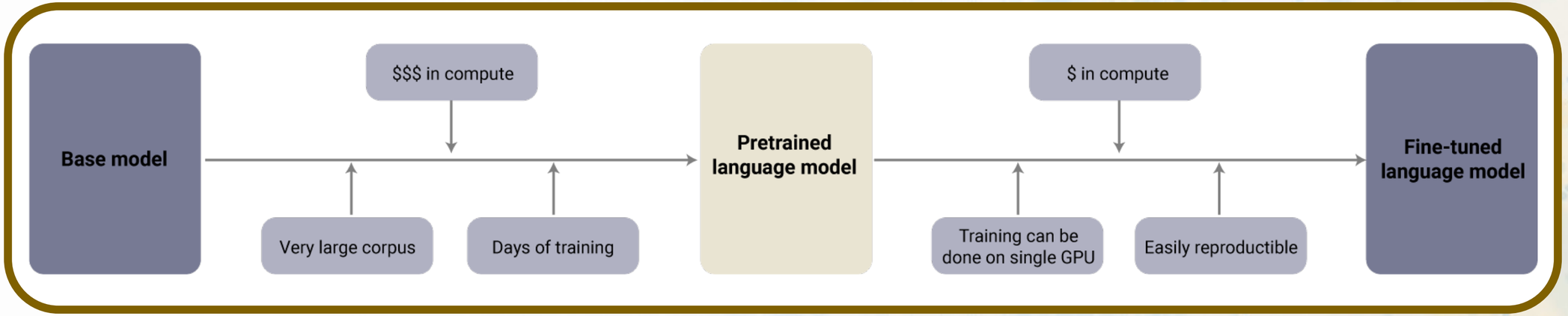
Why Were the Language Models Created?

It takes a long time to train Transformers on a large corpus of data

The training process is very lengthy and expensive, and creates a huge carbon footprint

Fine-tuning a pre-trained model requires much less data and time than the pre-training process

Pre-trained model is trained on a dataset similar to the fine-tuning dataset. Therefore, it has some knowledge about the dataset that we could use in fine-tuning process.



https://huggingface.co/course/chapter1/1

Transformers

A specific architecture originally developed for Machine Translation and is used for myriads of NLP tasks:

Sentiment analysis, Named Entity Recognition (NER), Part-of-Speech tagging (POS), text summarization, filling in the blanks, and so on.

https://huggingface.co/course/chapter1/1
Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).
Dorr, Bonnie J., Pamela W. Jordan, and John W. Benoit. "A survey of current paradigms in machine translation." *Advances in computers*. Vol. 49. Elsevier, 1999. 1-68.

NLP Problems Refresher

Let’s quickly go over some of these problems that Transformer models help with

- **Machine Translation:** "Machine translation (MT), … is, translation from one natural language into another by means of a computerized system …". *A survey of current paradigms in machine translation.1999*
- **Sentiment-Analysis:** identifying emotion of a sentence/phrase (positive, negative, neutral, ...)
- **Text Summarization:** summarizing a text into a shorter version
- **Fill in the blank:** This course will teach you all about ____ models. (mathematical, statistical, ...).
- **Named Entity Recognition (NER):** a very important method in NLP for detecting variety of entities in a text, e.g., people, places, date, etc.
- **Part-of-Speech Tagging (POS):** Assign a part-of-speech to a word in a sentence such as adjective, verb, etc

NLP Problems Refresher

Let’s quickly go over some of these problems that Transformer models help with

- **Machine Translation:** "Machine translation (MT), … is, translation from one natural language into another by means of a computerized system …". *A survey of current paradigms in machine translation.1999*
- **Sentiment-Analysis:** identifying emotion of a sentence/phrase (positive, negative, neutral, ...)
- **Text Summarization:** summarizing a text into a shorter version
- **Fill in the blank:** This course will teach you all about ____ models. (mathematical, statistical, ...).
- **Named Entity Recognition (NER):** a very important method in NLP for detecting variety of entities in a text, e.g., people, places, date, etc.
- **Part-of-Speech Tagging (POS):** Assign a part-of-speech to a word in a sentence such as adjective, verb, etc

Why Transformers?

LSTMs had been the state-of-the-art models for NLP tasks, but they have various problems:

- They are slow: items are fed to LSTMs one at a time (because they need the hidden state of the previous step to make predictions) and can’t be fed to the model in parallel.
- Usually need large memory
- Transfer learning doesn't work on them. So, training on a new set of labeled data is needed every time (not practical).

https://huggingface.co/course/chapter1/1
Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).
Dorr, Bonnie J., Pamela W. Jordan, and John W. Benoit. "A survey of current paradigms in machine translation." *Advances in computers*. Vol. 49. Elsevier, 1999. 1-68.