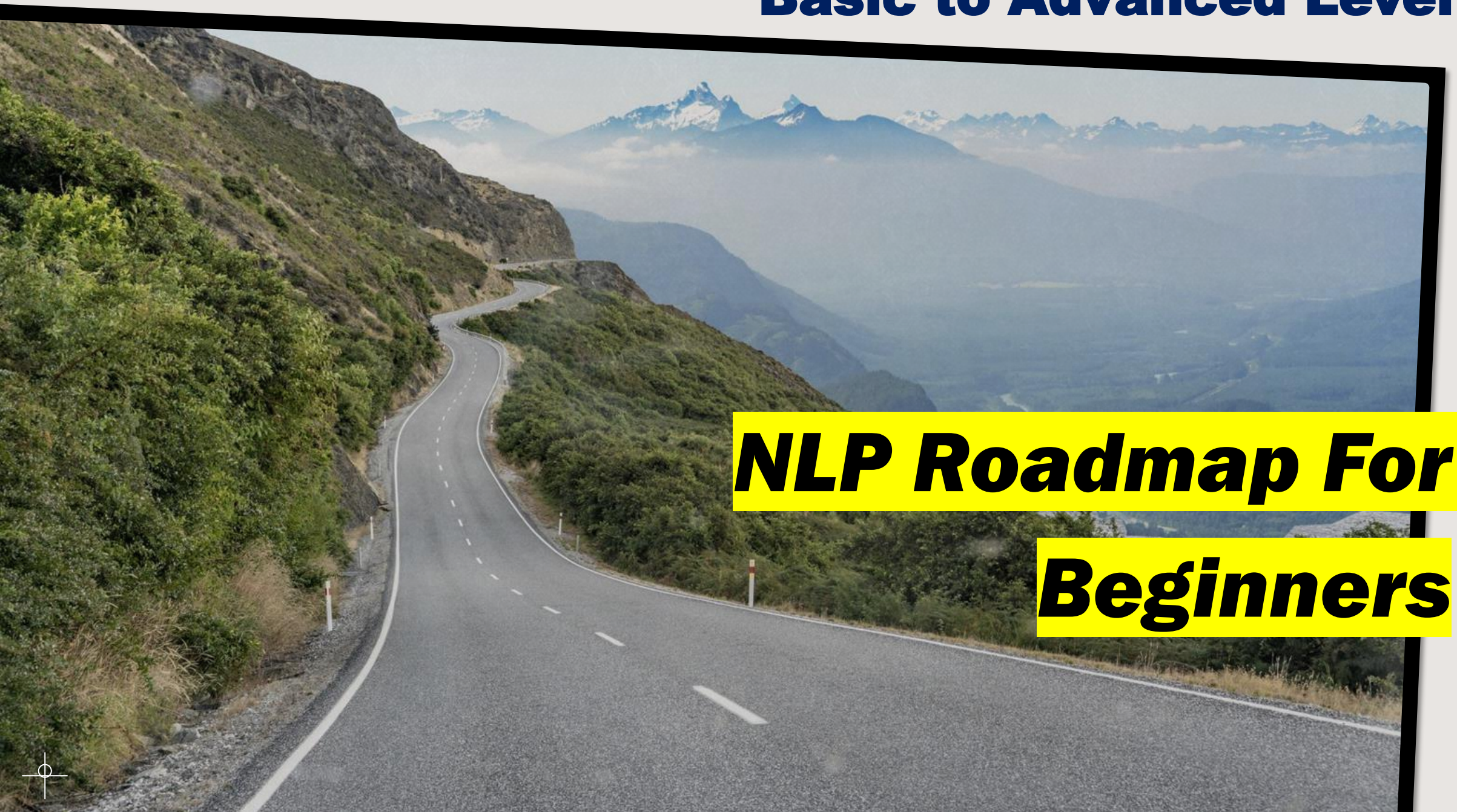
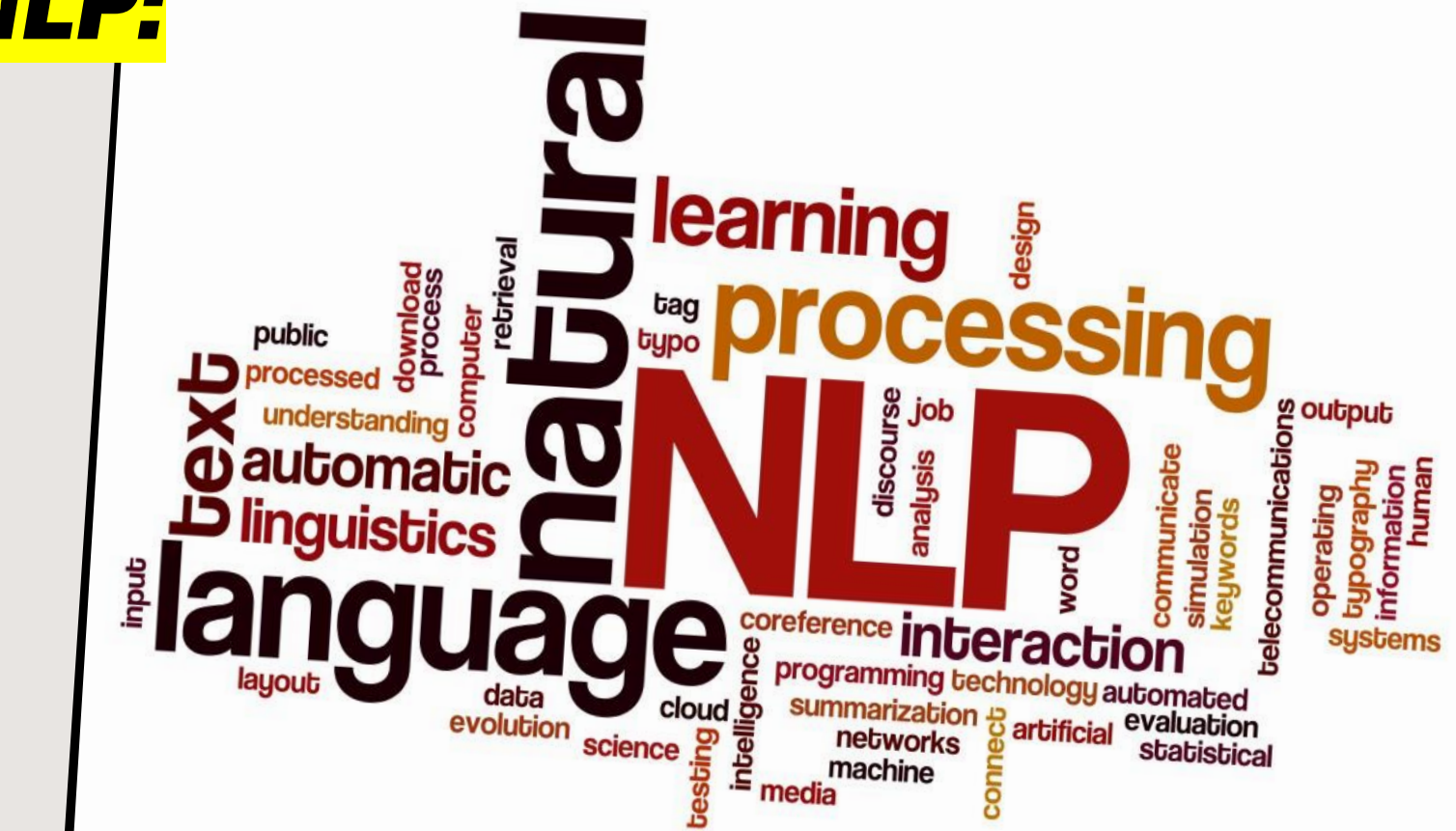
Basic to Advanced Level



NLP Roadmap For Beginners

Definition of NLP:

- Natural Language Processing (NLP) is a field of artificial intelligence that focuses on the interaction between computers and human language. It enables computers to understand, interpret, and generate human language in a way that is valuable and meaningful.



Applications of NLP:

Text Classification: Categorizing documents into predefined categories, such as spam detection or news topic classification.

Sentiment Analysis: Analyzing text to determine if it expresses positive, negative, or neutral emotions.

Language Translation: Systems like Google Translate that translate text from one language to another.

Chatbots: Virtual assistants like Siri and Alexa that understand and respond to spoken or typed language.

Text Summarization: Automatically generating concise summaries of longer texts.

Speech Recognition: Converting spoken language into text, as used in voice assistants and transcription services.

Types of Data in NLP:

Text: Written language in the form of documents, articles, emails, messages, and more.

Speech: Spoken language, often transcribed into textual form for analysis.

Multimodal Data: Combining text with other types of data, like images or videos for tasks such as image captioning or video transcription.

Structured Data: Text data combined with structured data, like databases or tables, for tasks such as information retrieval.

Tools and Technologies:

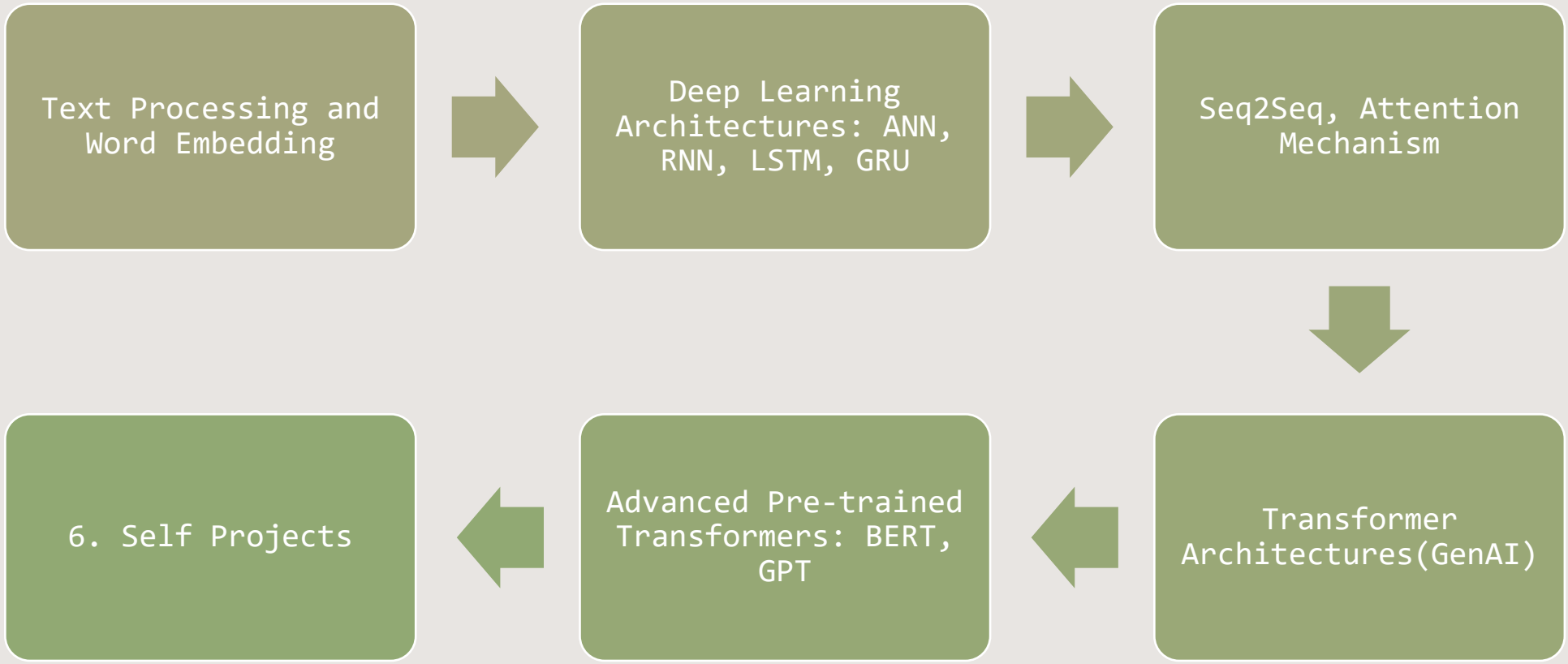
NLP relies on machine learning algorithms and neural networks to process and analyze language data.

Common programming languages for NLP: Python

Libraries: NLTK and spaCy

Deep Learning frameworks: TensorFlow and PyTorch.

Roadmap


Text Processing and Word Embedding
Deep Learning Architectures: ANN, RNN, LSTM, GRU
Seq2Seq, Attention Mechanism
Transformer Architectures(GenAI)
Advanced Pre-trained Transformers: BERT, GPT
6. Self Projects

Text Processing:

- Tokenization
- Stopwords
- Stemming
- Bag of Words (BoW)
- TF-IDF

Word Embedding:

- Word2vec
- GloVe

Neural Networks:

- Artificial Neural Networks (ANN)
- Recurrent Neural Networks (RNN)
- Long Short-Term Memory (LSTM)
- Gated Recurrent Unit (GRU)

Seq2Seq and Attention:

Sequence-to-Sequence (Seq2Seq) Models

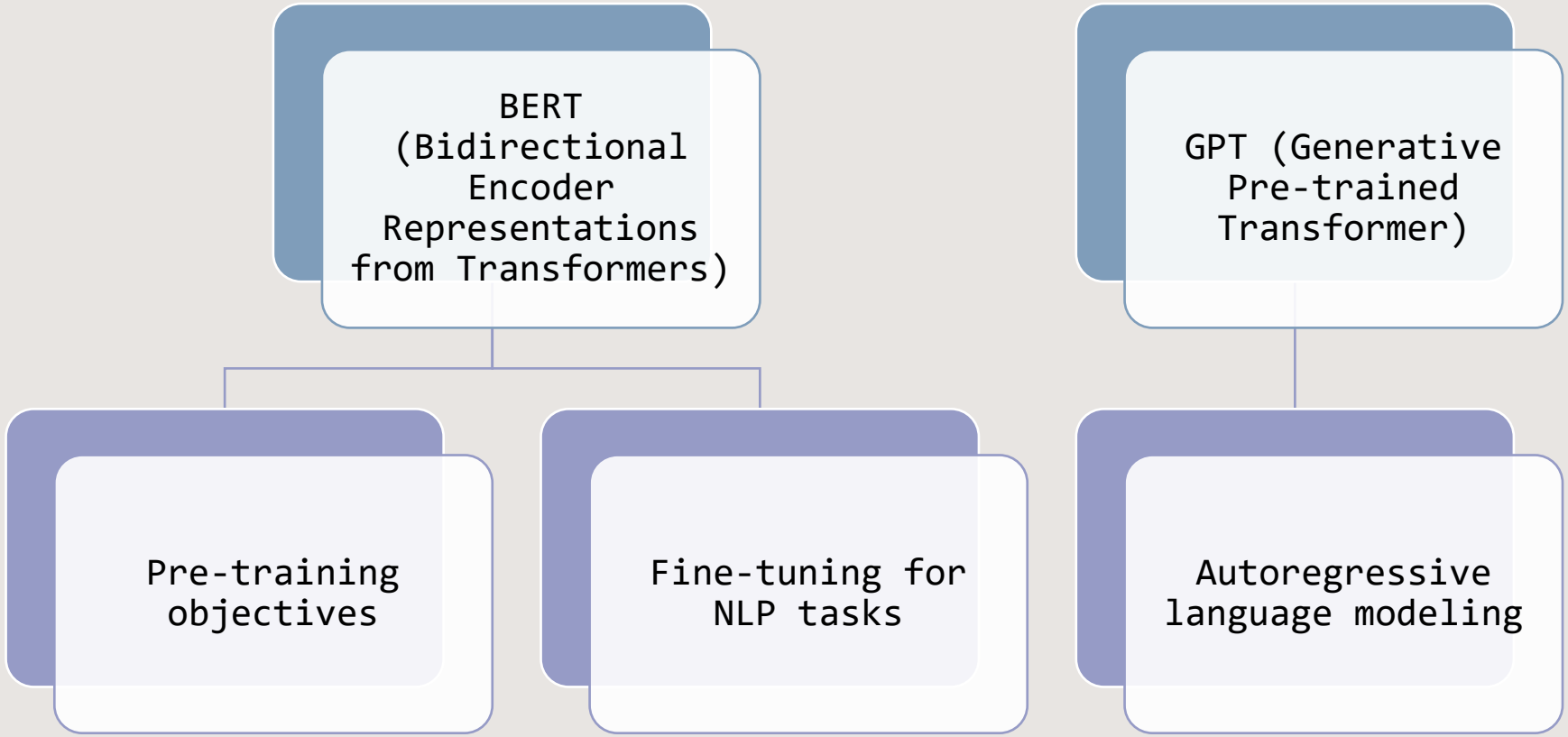
Attention Mechanism

Encoder-Decoder Architectures

Transformer:

- Transformer Architecture
- Self-Attention
- Multi-Head Attention
- ← Positional Encoding

BERT and GPT:


BERT (Bidirectional Encoder Representations from Transformers)
GPT (Generative Pre-trained Transformer)
Pre-training objectives
Fine-tuning for NLP tasks
Autoregressive language modeling

Practical Projects:

- Sentiment Analysis
- Text Summarization
- Named Entity Recognition
- Text Generation