The Transformer architecture, introduced in 2017 by Vaswani et al., is built entirely on attention mechanisms, removing recurrence entirely.

It consists of an encoder-decoder structure where both the encoder and decoder are made up of layers containing multi-head self-attention and feed-forward networks.

Transformers leverage positional encoding to retain order information in sequences.

The architecture allows for better parallelization and efficiency during training compared to RNNs.

Transformers form the backbone of modern NLP models such as BERT, GPT, T5, and more.

They have also been adapted for use in image processing, audio modeling, and multi-modal tasks.

The Transformer architecture, introduced in 2017 by Vaswani et al., is built entirely on attention mechanisms, removing recurrence entirely.

It consists of an encoder-decoder structure where both the encoder and decoder are made up of layers containing multi-head self-attention and feed-forward networks.

Transformers leverage positional encoding to retain order information in sequences.

The architecture allows for better parallelization and efficiency during training compared to RNNs.

Transformers form the backbone of modern NLP models such as BERT, GPT, T5, and more.

They have also been adapted for use in image processing, audio modeling, and multi-modal tasks.

The Transformer architecture, introduced in 2017 by Vaswani et al., is built entirely on attention mechanisms, removing recurrence entirely.

It consists of an encoder-decoder structure where both the encoder and decoder are made up of layers containing multi-head self-attention and feed-forward networks.

Transformers leverage positional encoding to retain order information in sequences.

The architecture allows for better parallelization and efficiency during training compared to RNNs.

Transformers form the backbone of modern NLP models such as BERT, GPT, T5, and more.

They have also been adapted for use in image processing, audio modeling, and multi-modal tasks.

The Transformer architecture, introduced in 2017 by Vaswani et al., is built entirely on attention mechanisms, removing recurrence entirely.

It consists of an encoder-decoder structure where both the encoder and decoder are made up of layers containing multi-head self-attention and feed-forward networks.

Transformers leverage positional encoding to retain order information in sequences.

The architecture allows for better parallelization and efficiency during training compared to RNNs.

Transformers form the backbone of modern NLP models such as BERT, GPT, T5, and more.

They have also been adapted for use in image processing, audio modeling, and multi-modal tasks.

The Transformer architecture, introduced in 2017 by Vaswani et al., is built entirely on attention mechanisms, removing recurrence entirely.

It consists of an encoder-decoder structure where both the encoder and decoder are made up of layers containing multi-head self-attention and feed-forward networks.

Transformers leverage positional encoding to retain order information in sequences.

The architecture allows for better parallelization and efficiency during training compared to RNNs.

Transformers form the backbone of modern NLP models such as BERT, GPT, T5, and more.

They have also been adapted for use in image processing, audio modeling, and multi-modal tasks.

The Transformer architecture, introduced in 2017 by Vaswani et al., is built entirely on attention mechanisms, removing recurrence entirely.

It consists of an encoder-decoder structure where both the encoder and decoder are made up of layers containing multi-head self-attention and feed-forward networks.

Transformers leverage positional encoding to retain order information in sequences.

The architecture allows for better parallelization and efficiency during training compared to RNNs.

Transformers form the backbone of modern NLP models such as BERT, GPT, T5, and more.

They have also been adapted for use in image processing, audio modeling, and multi-modal tasks.

The Transformer architecture, introduced in 2017 by Vaswani et al., is built entirely on attention mechanisms, removing recurrence entirely.

It consists of an encoder-decoder structure where both the encoder and decoder are made up of layers containing multi-head self-attention and feed-forward networks.

Transformers leverage positional encoding to retain order information in sequences.

The architecture allows for better parallelization and efficiency during training compared to RNNs.

Transformers form the backbone of modern NLP models such as BERT, GPT, T5, and more.

They have also been adapted for use in image processing, audio modeling, and multi-modal tasks.

The Transformer architecture, introduced in 2017 by Vaswani et al., is built entirely on attention mechanisms, removing recurrence entirely.

It consists of an encoder-decoder structure where both the encoder and decoder are made up of layers containing multi-head self-attention and feed-forward networks.

Transformers leverage positional encoding to retain order information in sequences.

The architecture allows for better parallelization and efficiency during training compared to RNNs.

Transformers form the backbone of modern NLP models such as BERT, GPT, T5, and more.

They have also been adapted for use in image processing, audio modeling, and multi-modal tasks.

The Transformer architecture, introduced in 2017 by Vaswani et al., is built entirely on attention mechanisms, removing recurrence entirely.

It consists of an encoder-decoder structure where both the encoder and decoder are made up of layers containing multi-head self-attention and feed-forward networks.

Transformers leverage positional encoding to retain order information in sequences.

The architecture allows for better parallelization and efficiency during training compared to RNNs.

Transformers form the backbone of modern NLP models such as BERT, GPT, T5, and more.

They have also been adapted for use in image processing, audio modeling, and multi-modal tasks.

The Transformer architecture, introduced in 2017 by Vaswani et al., is built entirely on attention mechanisms, removing recurrence entirely.

It consists of an encoder-decoder structure where both the encoder and decoder are made up of layers containing multi-head self-attention and feed-forward networks.

Transformers leverage positional encoding to retain order information in sequences.

The architecture allows for better parallelization and efficiency during training compared to RNNs.

Transformers form the backbone of modern NLP models such as BERT, GPT, T5, and more.

They have also been adapted for use in image processing, audio modeling, and multi-modal tasks.

The Transformer architecture, introduced in 2017 by Vaswani et al., is built entirely on attention mechanisms, removing recurrence entirely.

It consists of an encoder-decoder structure where both the encoder and decoder are made up of layers containing multi-head self-attention and feed-forward networks.

Transformers leverage positional encoding to retain order information in sequences.

The architecture allows for better parallelization and efficiency during training compared to RNNs.

Transformers form the backbone of modern NLP models such as BERT, GPT, T5, and more.

They have also been adapted for use in image processing, audio modeling, and multi-modal tasks.

The Transformer architecture, introduced in 2017 by Vaswani et al., is built entirely on attention mechanisms, removing recurrence entirely.

It consists of an encoder-decoder structure where both the encoder and decoder are made up of layers containing multi-head self-attention and feed-forward networks.

Transformers leverage positional encoding to retain order information in sequences.

The architecture allows for better parallelization and efficiency during training compared to RNNs.

Transformers form the backbone of modern NLP models such as BERT, GPT, T5, and more.

They have also been adapted for use in image processing, audio modeling, and multi-modal tasks.

The Transformer architecture, introduced in 2017 by Vaswani et al., is built entirely on attention mechanisms, removing recurrence entirely.

It consists of an encoder-decoder structure where both the encoder and decoder are made up of layers containing multi-head self-attention and feed-forward networks.

Transformers leverage positional encoding to retain order information in sequences.

The architecture allows for better parallelization and efficiency during training compared to RNNs.

Transformers form the backbone of modern NLP models such as BERT, GPT, T5, and more.

They have also been adapted for use in image processing, audio modeling, and multi-modal tasks.

The Transformer architecture, introduced in 2017 by Vaswani et al., is built entirely on attention mechanisms, removing recurrence entirely.

It consists of an encoder-decoder structure where both the encoder and decoder are made up of layers containing multi-head self-attention and feed-forward networks.

Transformers leverage positional encoding to retain order information in sequences.

The architecture allows for better parallelization and efficiency during training compared to RNNs.

Transformers form the backbone of modern NLP models such as BERT, GPT, T5, and more.

They have also been adapted for use in image processing, audio modeling, and multi-modal tasks.

The Transformer architecture, introduced in 2017 by Vaswani et al., is built entirely on attention mechanisms, removing recurrence entirely.

It consists of an encoder-decoder structure where both the encoder and decoder are made up of layers containing multi-head self-attention and feed-forward networks.

Transformers leverage positional encoding to retain order information in sequences.

The architecture allows for better parallelization and efficiency during training compared to RNNs.

Transformers form the backbone of modern NLP models such as BERT, GPT, T5, and more.

They have also been adapted for use in image processing, audio modeling, and multi-modal tasks.

The Transformer architecture, introduced in 2017 by Vaswani et al., is built entirely on attention mechanisms, removing recurrence entirely.

It consists of an encoder-decoder structure where both the encoder and decoder are made up of layers containing multi-head self-attention and feed-forward networks.

Transformers leverage positional encoding to retain order information in sequences.

The architecture allows for better parallelization and efficiency during training compared to RNNs.

Transformers form the backbone of modern NLP models such as BERT, GPT, T5, and more.

They have also been adapted for use in image processing, audio modeling, and multi-modal tasks.

The Transformer architecture, introduced in 2017 by Vaswani et al., is built entirely on attention mechanisms, removing recurrence entirely.

It consists of an encoder-decoder structure where both the encoder and decoder are made up of layers containing multi-head self-attention and feed-forward networks.

Transformers leverage positional encoding to retain order information in sequences.

The architecture allows for better parallelization and efficiency during training compared to RNNs.

Transformers form the backbone of modern NLP models such as BERT, GPT, T5, and more.

They have also been adapted for use in image processing, audio modeling, and multi-modal tasks.

The Transformer architecture, introduced in 2017 by Vaswani et al., is built entirely on attention mechanisms, removing recurrence entirely.

It consists of an encoder-decoder structure where both the encoder and decoder are made up of layers containing multi-head self-attention and feed-forward networks.

Transformers leverage positional encoding to retain order information in sequences.

The architecture allows for better parallelization and efficiency during training compared to RNNs.

Transformers form the backbone of modern NLP models such as BERT, GPT, T5, and more.

They have also been adapted for use in image processing, audio modeling, and multi-modal tasks.

The Transformer architecture, introduced in 2017 by Vaswani et al., is built entirely on attention mechanisms, removing recurrence entirely.

It consists of an encoder-decoder structure where both the encoder and decoder are made up of layers containing multi-head self-attention and feed-forward networks.

Transformers leverage positional encoding to retain order information in sequences.

The architecture allows for better parallelization and efficiency during training compared to RNNs.

Transformers form the backbone of modern NLP models such as BERT, GPT, T5, and more.

They have also been adapted for use in image processing, audio modeling, and multi-modal tasks.

The Transformer architecture, introduced in 2017 by Vaswani et al., is built entirely on attention mechanisms, removing recurrence entirely.

It consists of an encoder-decoder structure where both the encoder and decoder are made up of layers containing multi-head self-attention and feed-forward networks.

Transformers leverage positional encoding to retain order information in sequences.

The architecture allows for better parallelization and efficiency during training compared to RNNs.

Transformers form the backbone of modern NLP models such as BERT, GPT, T5, and more.

They have also been adapted for use in image processing, audio modeling, and multi-modal tasks.