Transformers and deep learning are related concepts, but they have some key differences.

1. Model Architecture: Deep learning typically refers to neural networks with many layers, also known as deep neural networks. These networks are designed to automatically learn features from data through multiple layers of nonlinear transformations. Transformers, on the other hand, are a specific type of deep learning model architecture that was introduced in the paper "Attention is All You Need" by Vaswani et al. in 2017. Transformers are primarily used for sequence-to-sequence tasks, such as machine translation, and are based on a self-attention mechanism, which allows the model to weigh the importance of different words in a sentence when making predictions.
2. Attention Mechanism: Transformers are known for their self-attention mechanism, which allows the model to attend to different parts of the input sequence when making predictions. This allows transformers to capture long-range dependencies and relationships between words in a sentence, making them well-suited for tasks that require modeling of contextual information. Deep learning models, on the other hand, typically rely on fixed-size convolutional or recurrent filters that scan the input sequentially or locally, without the ability to attend to different parts of the input sequence simultaneously.
3. Parallelism: Transformers are highly parallelizable, as the self-attention mechanism allows for parallel processing of input sequences, making them efficient for training on modern hardware with multiple GPUs. Deep learning models, on the other hand, may have limitations in terms of parallelism due to their sequential nature, which can affect their training and inference efficiency.
4. Applications: Deep learning models have been widely used in a variety of tasks, such as image recognition, speech recognition, natural language processing, and playing games, among others. Transformers, on the other hand, have been particularly successful in tasks that involve processing sequences, such as machine translation, language modeling, text generation, and sentiment analysis, among others.
5. Training Data: Both deep learning models and transformers require large amounts of training data to effectively learn from the data and make accurate predictions. However, transformers may require even larger amounts of data due to their capacity for capturing long-range dependencies and modeling complex relationships between words in a sentence.