

# Transformerアーキテクチャの代表的な種類

- エンコーダ・デコーダ (Encoder-decoder)
  - ❖ オリジナルのTransformerと同様に、エンコーダ・デコーダの2つのブロックで構成される
  - ❖ エンコーダーは、入力されたトークンの列をから入力と同じ長さのベクトル列を出力する
  - ❖ デコーダは、エンコーダの出力内容からターゲットとなるシーケンスをトークンごとに自己回帰的に予測する
  - ❖ T5やBARTなどで採用されている

# Transformerアーキテクチャの代表的な種類

- 因果的デコーダ (Causal Decoder)
  - ❖ テキスト列を自己回帰的に予測するようにトレーニングされたデコーダのみのアーキテクチャ
  - ❖ 「因果的」とは、モデルが左の文脈（次のステップの予測）だけに依存することを意味する
  - ❖ 一方向のアテンション・マスクが組み込まれており、各入力トークンは過去のトークンと自分自身にのみアテンションできる
  - ❖ GPTシリーズやBLOOMなどで採用されている