

# 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

# Transformerアーキテクチャの代表的な種類

- プレフィックスデコーダ (Prefix Decoder)
  - ❖ デコーダオンリーモデルが入力テキストのより豊かな非因果的表現を構築できるように改良されたアーキテクチャ
  - ❖ 入力シーケンスが非因果的マスクを持つように（すなわち、過去のトークンに制限されない）、アテンションマスクが変更されている
  - ❖ prefixトークンに対する双方向の注意と、生成されたトークンに対する一方向の注意を可能にする
  - ❖ U-PaLMやGLM-130Bなどで採用されている