Core Idea: A Hybrid of BERT and GPT
BART is a Transformer-based encoder-decoder model that can be seen as a blend of BERT and GPT:

Like BERT → BART has a bidirectional encoder, meaning it can attend to both left and right context.
Like GPT → BART has a left-to-right autoregressive decoder, making it suitable for text generation.
However, unlike BERT (which is designed for token-level classification) and GPT (which is purely generative), BART is trained as a denoising autoencoder, meaning it learns to reconstruct corrupted text inputs. This makes it highly effective for tasks like summarization, translation, and text generation.

--> Pretraining with Denoising Autoencoding

How does BART learn?
BART is pretrained on a corrupted input text and learns to reconstruct the original text. This forces the model to develop strong representations of language, making it more robust.

Unlike BERT (which uses masked language modeling), BART supports various types of noising functions:

Key Noising Functions (Corruptions)
Token Masking – Random tokens are replaced with <mask> (like BERT).
Token Deletion – Some tokens are completely removed (no placeholder).
Sentence Permutation – The order of sentences is shuffled.
Text Infilling – Spans of text are replaced with a single <mask> token (similar to T5).
Document Rotation – The text is rotated by shifting a random token to the beginning.
These corruptions force the model to develop deep contextual understanding by learning to predict missing or scrambled text.

--> Architecture: Standard Transformer Encoder-Decoder

BART’s architecture is a standard Transformer-based seq2seq model:

Encoder (like BERT) → Fully bidirectional attention.
Decoder (like GPT) → Left-to-right autoregressive generation.

--> Why is BART Effective?

The bidirectional encoder captures rich contextual meaning (like BERT).
The autoregressive decoder enables coherent text generation (like GPT).
Pretraining with multiple noise functions makes BART more robust than BERT or GPT alone.

--> Fine-Tuning for Downstream Tasks
After pretraining, BART can be fine-tuned on various NLP tasks:

(1) Text Generation (Summarization, Translation)
Since BART’s decoder is trained autoregressively, it excels at:

Summarization: Given a long document, generate a concise summary.
Machine Translation: Convert text from one language to another.

(2) Text Classification (Sentiment, Natural Language Inference)
Even though BART is a seq2seq model, it can be fine-tuned for classification by:

Feeding the input into the encoder.
Using the decoder to output a class token.

(3) Question Answering (QA)

BART can also be used for extractive or abstractive QA:

Extractive QA (SQuAD-like tasks) → It can predict spans of answers from a given context.
Abstractive QA → It can generate natural-sounding answers.

Hybrid of BERT & GPT → It combines bidirectional understanding (BERT) with autoregressive generation (GPT).
Denoising Autoencoder → Learns to reconstruct text from noisy inputs, making it robust.
Strong on Summarization, QA, and Translation → BART is particularly good for text generation tasks.
Flexible Architecture → Can be fine-tuned for classification, QA, and more.
Pretrained on Diverse Noising Strategies → Improves generalization across multiple NLP benchmarks.