# Attention Mechanism

Part 5/7: Architecture

## Self-Attention

Q | K | V

Models long-range dependencies in image. Applied at 16×16, 32×32 resolutions.

## Cross-Attention

Text Q → Image K,V

For text conditioning. Text→Image attention enables controllable generation.

## Key Properties

**Multi-Head**
Multiple patterns

**Spatial Attention**
Different locations

**Complexity**
$O(n^2)$

**Mechanism**
Standard transformer

**Location**
Low resolution

**Dependencies**
Long-range

✨ Impact: Crucial for coherent, high-quality generation