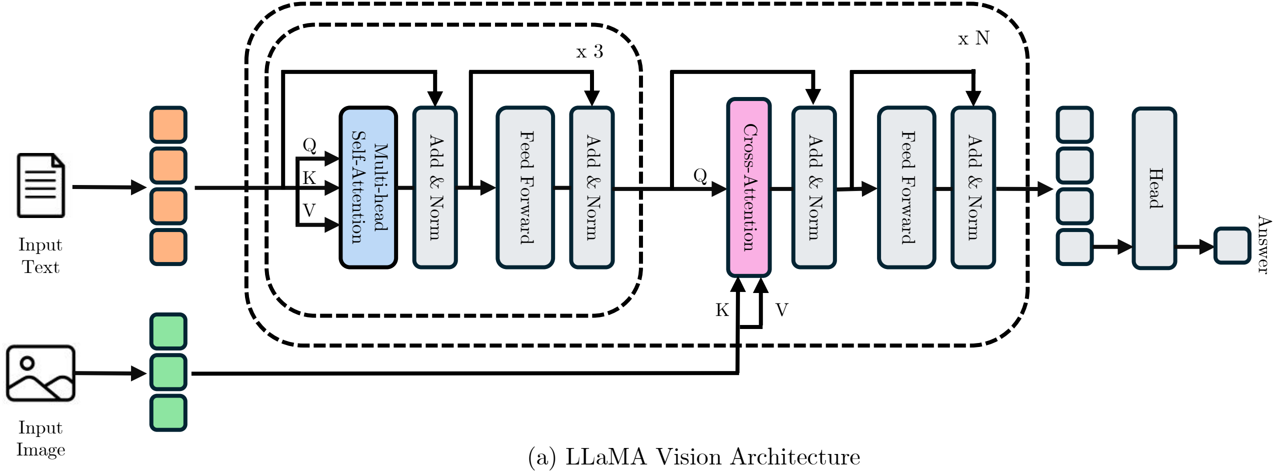

(a) LLaMA Vision Architecture

0	-inf	-inf	-inf
0	0	-inf	-inf
0	0	0	-inf
0	0	0	0

Self-attention mask

0	0	0
0	0	0
0	0	0
0	0	0

Cross-attention mask

(b) Default Attention Masks

0	-inf	-inf	-inf
0	0	-inf	-inf
0	0	0	-inf
0	0	0	-inf

Last ↛ Last

0	-inf	-inf	-inf
0	0	-inf	-inf
0	0	0	-inf
-inf	-inf	-inf	0

Question ↛ Last

0	0	0
0	0	0
0	0	0
-inf	-inf	-inf

Image ↛ Last

-inf	-inf	-inf
-inf	-inf	-inf
-inf	-inf	-inf
0	0	0

Image ↛ Question

(c) Attention Masking Mechanism to analyze different information flows