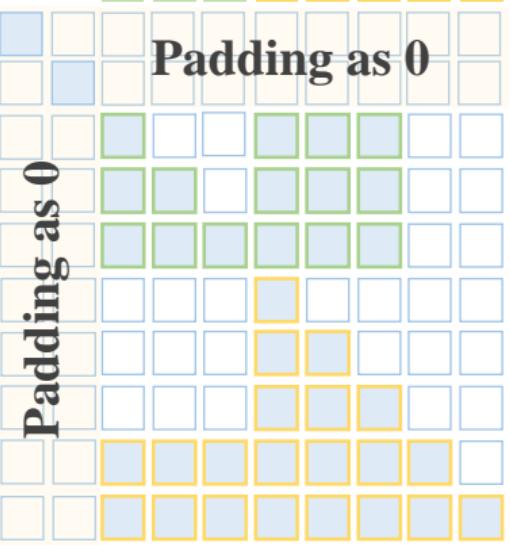


3 4 5 0 1 2 6 7



Padding as 0

= 0

Padding = 0
when $bs > 1$

Text mask = 1

Visual mask = 1

unified causal mask

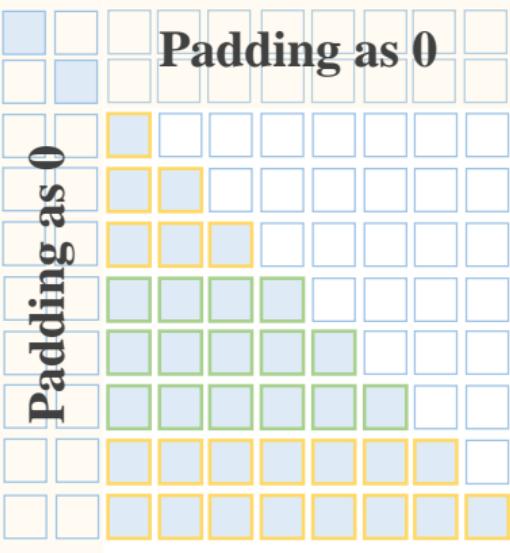
7 6 5 4 3 2 1 0

7 6 5 4 3 2 1 0

Unified
token
index

naive causal mask

7 6 5 4 3 2 1 0



Padding as 0

AIFS