Visual Tokens

Prompt Tokens

Generated Text Tokens

Image Encoder

Input Image

Scan this image sequentially based on added symbols (@, #, $) and describe it.

User Prompt

Row @: triangle
Row #: circle
Row $: square

Generated Response

@ partition

partition

$ partition

Row @: : triangle \n Row # : circle \n Row $: square <EOS>