

# Number of parameters

203M

424M

1B

4B

>4B

Perception Test VQA accuracy

60

55

50

45

40

35

20

25

30

35

40

45

50

EgoSchema VQA accuracy

SeViLA

LongViViT

ImageViT

Bard + ImageViT

ShortViViT

Bard + ShortViViT

MC-ViT-B

Bard + PALI

MC-ViT-L