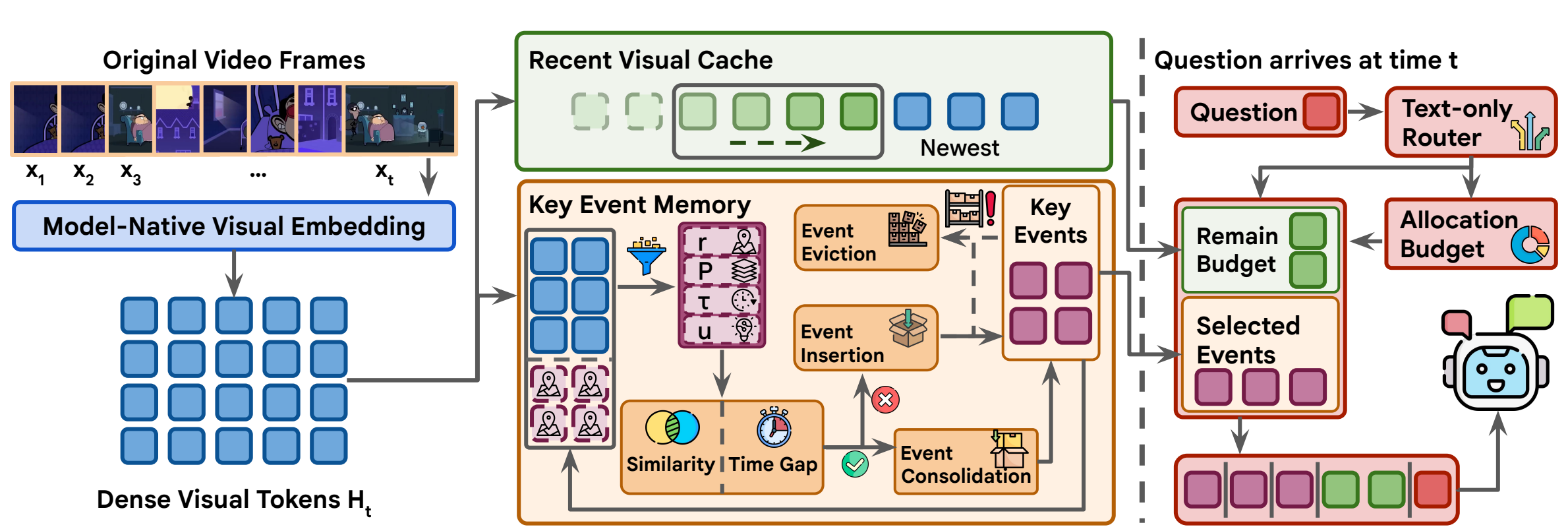

Original Video Frames
x1
x2
x3
...
xt
Model-Native Visual Embedding
Dense Visual Tokens Ht
Recent Visual Cache
Newest
Key Event Memory
r
P
τ
u
Similarity
Time Gap
Event Eviction
Event Insertion
Event Consolidation
Key Events
Question arrives at time t
Question
Text-only Router
Allocation Budget
Remain Budget
Selected Events