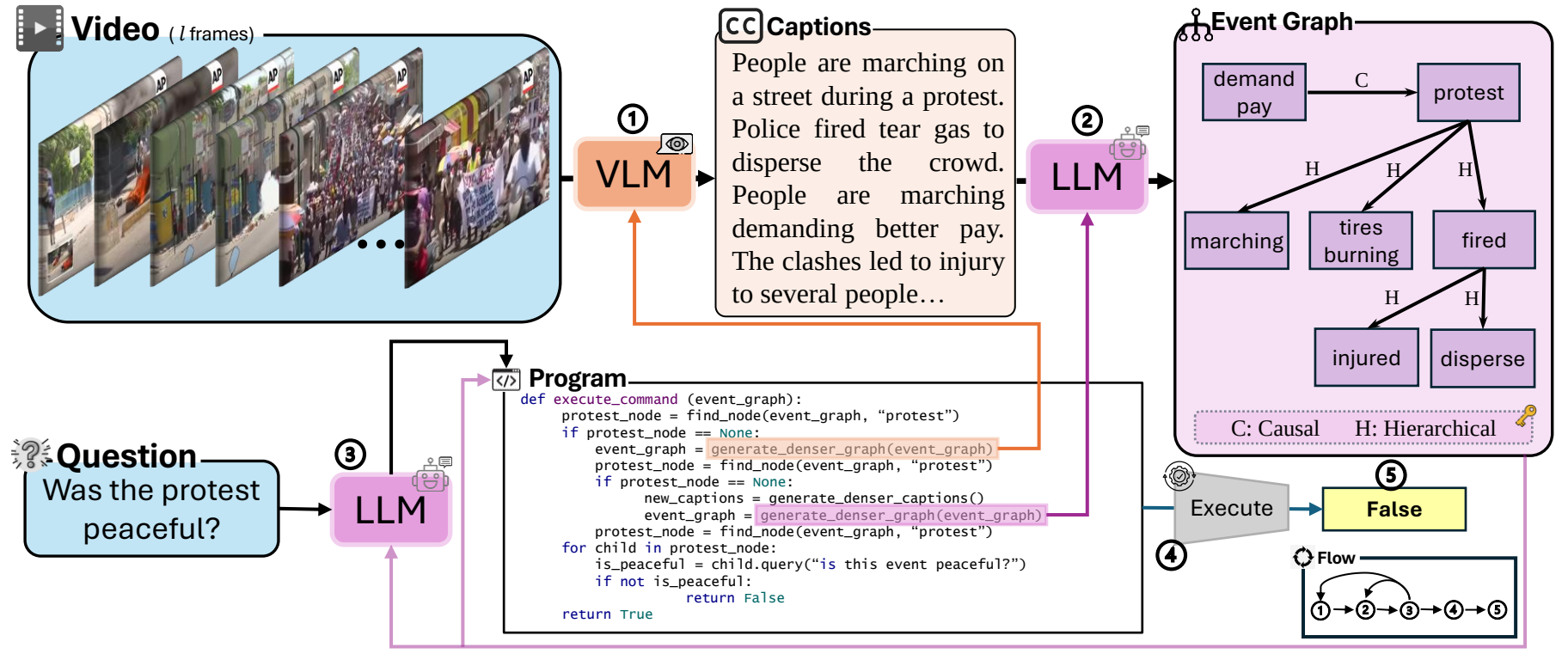
Video (l frames)
①
VLM
Captions
People are marching on a street during a protest. Police fired tear gas to disperse the crowd. People are marching demanding better pay. The clashes led to injury to several people…
②
LLM
Event Graph
demand pay
C
protest
H
H
H
marching
tires burning
fired
H
H
injured
disperse
C: Causal
H: Hierarchical
Question
Was the protest peaceful?
③
LLM
Program
def execute_command (event_graph):
protest_node = find_node(event_graph, "protest")
if protest_node == None:
event_graph = generate_denser_graph(event_graph)
protest_node = find_node(event_graph, "protest")
if protest_node == None:
new_captions = generate_denser_captions()
event_graph = generate_denser_graph(event_graph)
protest_node = find_node(event_graph, "protest")
for child in protest_node:
is_peaceful = child.query("is this event peaceful?")
if not is_peaceful:
return False
return True
④
Execute
⑤
False
Flow
① ② ③ ④ ⑤