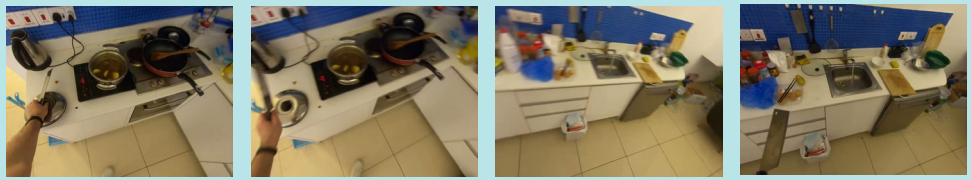
Streaming Video Understanding

Action Prediction

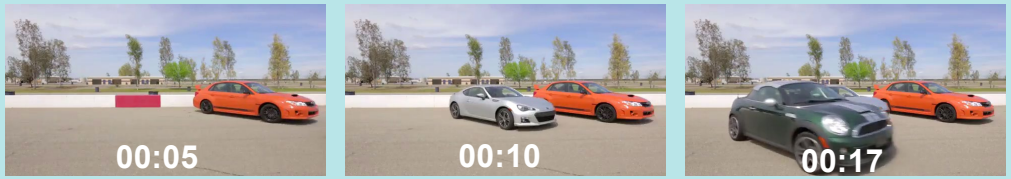
I am tasked with Peel, chop and cook the second potato. What is next step?



A: pick up potato

Dynamic State Grounding

How many cars have shown up?



A: 1 A: 2 A: 3

Multi-Turn Dependency Reasoning

Q: Who is organizing the backpack?
A: A woman

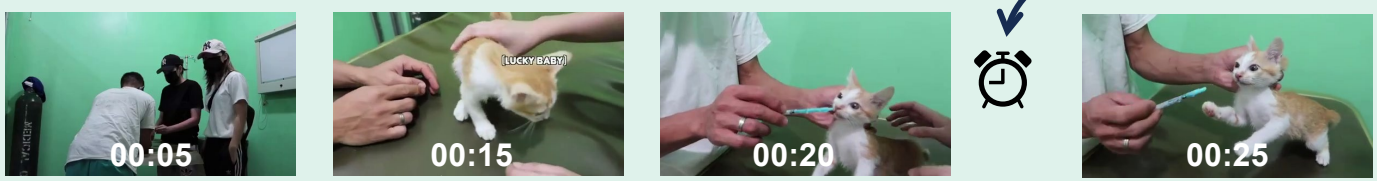
Q: Who is <a woman> talk to?
A: A man

Q: What is <a man> doing?
A: Sitting on the bed.

Proactive Reasoning

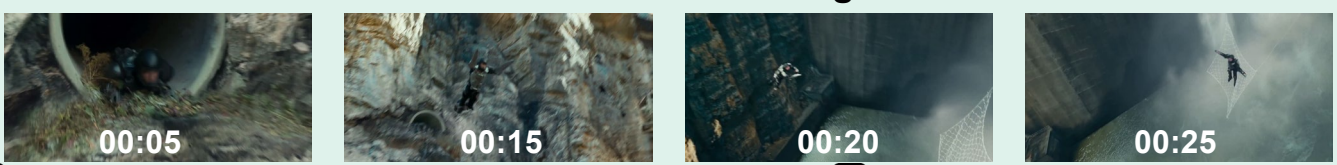
Proactive Alerting

Inform me when a cat is getting a shot.



A: Informed

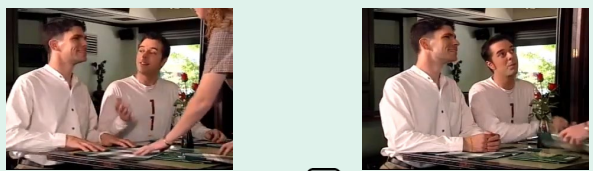
Proactive Turn-Taking



What happens to the man in black clothes after he stands on the cliff?
A: He is caught by a net laid by a helicopter.

Oh well, it is what it is. No point in overthinking it.
A: <Silence>

Speaker Identification



This is Bob

This is Wanita

Who wants to serve drink?
A: Wanita