(a): Meta-Learning Capabilities

Qwen3 FT (Zero-Shot):
A little girl in an Elmo costume holding a guitar.

Qwen3 FT (Few-Shot):
A young blond child in a red Elmo suit is playing on a red Elmo guitar.

Ground Truth:
A child in a Elmo suit is playing the guitar.

Success: Few-shot prompts recovered details ('blond', 'red guitar').

(b): Contextual Interference

Llama 3.2 FT (Zero-Shot):
Two young men with brown hair... sit at a table surrounded by cords...

Llama 3.2 FT (Few-Shot):
Two men... one with a cigarette... one with a pencil in his mouth... playing a game.

Ground Truth:
Two men looking at little white laptops and holding a guitar.

Failure: Few-shot prompts caused hallucinations ('cigarette', 'pencil').

(c): Fine-Grained Limitation

Llama 3.2 FT (Zero-Shot):
A man in a white shirt walks toward the camera as people walk away...

Qwen3 FT (Few-Shot):
An elderly man in a white shirt is walking in a public area while other people walk behind him.

Ground Truth:
An old man walking with a folder in his hand.

Failure: Both models missed the small object ('folder') mentioned in GT.