

(1) Base Model w/o Reasoning Demos



System: [Task Description]



User: [Support Image] [Support Question]



Assistant: [Support Answer (GT)]



User: [Query Image] [Query Question]



Assistant: [Query Answer (Generated)]

(2) Reasoning Model w/o Reasoning Demos



System: [Task Description]



User: [Support Image] [Support Question]



Assistant: [Support Answer (GT)]



User: [Query Image] [Query Question]



Assistant: [Query Reasoning (Generated)]

(3) Base Model w/ Reasoning Demos



System: [Task Description]



User: [Support Image] [Support Question]

[Support Reasoning (GT)]



Assistant: [Support Answer (GT)]



User : [Query Image] [Query Question]



Assistant: [Query Answer (Generated)]

(4) Reasoning Model w/ Reasoning Demos



System: [Task Description]



User: [Support Image] [Support Question]



Assistant: [Support Reasoning (Pseudo / Gold / GT)]



User: [Query Image] [Query Question]



Assistant: [Query Reasoning (Generated)]