```
- Use the exact chat_id numbers from the conversation turns

## OUTPUT FORMAT
You will output exactly ONE JSON object matching this schema:
{
  "question": string, "answer": string, "difficulty": "hard", "reasoning_type": [Some categories]
  "sessions_required": integer, "conversation_references": [string,...], "reasoning_steps": [string,...], "
      source_chat_ids": [integer,...]
}

## IMPORTANT REQUIREMENTS
1. **Complex multi-session dependency**: Question must require sophisticated information synthesis from 4+
      conversation sessions
2. **Question source**: The question and answer to it MUST BE based on information from USER messages in
      CONVERSATION TURNS. You MUST NOT generate questions about assistant responses or suggestions.
3. **User information only**: Only create questions that test details the user provided, not assistant advice
      or recommendations.
5. **Advanced reasoning path**: Provide complex reasoning steps that require multi-hop thinking
6. **Precise answer**: Give exact answer that demonstrates complex analysis
7. **Session references**: Note which sessions contain relevant information
8. **High complexity**: Ensure question requires advanced multi-session reasoning and sophisticated synthesis

Generate ONE high-quality hard multi-session reasoning question.

NOTE: Only output the JSON object without any explanation before or after.
```

Listing 11: Multi-hop reasoning probing question generation prompt

```
You are tasked with generating a probing question to test knowledge update capabilities of LLMs. You will be
      given two related bullet points (original information and updated information) and the corresponding
      multi-turn dialogs between a user and assistant that incorporate both pieces of information across
      different conversation sessions.

Your task is to create ONE question that asks about the current/updated state of information, testing whether
      the LLM correctly recalls the most recent version rather than outdated information.

## INPUT DATA
- **BULLET POINTS**: <bullet_points>
- **CONVERSATION TURNS**: <conversation_turns>

## CRITICAL REQUIREMENT: NO SPECIFIC CONTEXT HINTS
- MUST NOT GIVE ANY INFORMATION RELATED TO OLD AND UPDATED INFORMATION/FACTS OR ANY HINTS THAT THERE IS UPDATE
      AT ALL
_ DO NOT use words like: currently, now, ... that shows update of information
- The question MUST NOT include specific dates, times, locations, or detailed circumstances
- Do NOT reference specific events, phases, or instances that would hint at which version to recall
- Ask about the general current state, not specific occurrences

## FACTUAL UPDATE IDENTIFICATION
Before creating the question:
1. Identify the EXACT fact that was updated in the "Information Update" bullet
2. Determine what the original fact was vs. the updated fact
3. Create a question that tests recall of the updated fact specifically
4. Ensure question asks for the factual detail, not procedures or implications

## QUESTION GENERATION GUIDELINES
Focus on creating questions that:
- **Ask about current state**: Question the most recent/updated version of information
- **Test update retention**: Whether LLM remembers the latest information, not the original
- **Avoid mentioning changes**: Don't explicitly ask "how did X change" - just ask about current state
- **Target updated facts**: Focus on information that was specifically updated/changed

## QUESTION TYPES TO GENERATE
1. **Current State Query** 2. **Latest Status** 3. **Updated Decision** 4. **Final Information** 5. **Recent
      Details**

## QUESTION LANGUAGE REQUIREMENTS
- Write questions as if the USER is asking them naturally
- **If testing information from USER messages**: Use first person ("I", "my", "me") in question -> Answer uses
      ("you", "your")
  - Example: "How did I decide on the location?" -> "You decided on the location because..."
- **If testing information from ASSISTANT messages**: Use second person ("you", "your") in question -> Answer
      uses ("I", "my")
  - Example: "What steps did you suggest for handling this?" -> "I suggested doing..."
- Avoid phrases like "according to the conversation", "based on what was discussed", "from our chat history"
- Make questions sound conversational and natural
- Questions should flow naturally as if continuing the conversation

## CHAT ID TRACKING REQUIREMENT
- You MUST identify which specific chat_id(s) contain the original and updated information
- List the chat_id with the original information and the chat_id with the updated information
- NOTE: If the answer is spread out between multiple chat_ids, group them in one list
- NOTE: DO NOT INCLUDE chat_ids in the answer
- Use the exact chat_id numbers from the conversation turns

## OUTPUT FORMAT
Return your analysis in this exact JSON format:
{
    "question": "", "answer": "", "difficulty": "moderate", "update_type": "", "tests_retention_of": "", "
        conversation_references": "",
    "potential_confusion": "", "source_chat_ids": {"original_info": [, ], "updated_info": [, ]}
```