```
- 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
- If answer spans multiple chats, include all relevant chat_ids - Use the exact chat_id numbers from the
      conversation turns

## DIFFICULTY LEVEL: HARD
- **Hard**: Requires synthesizing multiple details from different parts of conversation
- Force integration of information scattered across multiple conversation turns
- Test ability to connect related facts from different conversation contexts
- Require deep understanding and synthesis rather than simple recall

## OUTPUT FORMAT
Return your analysis in this exact JSON format:
{
    "question": [], "answer": [], "difficulty": "hard", "question_type": # one of: [] "conversation_reference
         ": "", "key_facts_tested": "",
    "extraction_challenge": "", "source_chat_ids": [X, Y, ...]
}

## IMPORTANT REQUIREMENTS
1. **Indirect questioning**: Ask about context rather than direct facts
2. **Question source flexibility**: Questions can be based on information from EITHER user messages OR
      assistant messages
3. **Perspective matching**: Question perspective must match the source of information:
   - **User info** -> "I/my/me" question -> "you/your" answer
   - **Assistant info** -> "you/your" question -> "I/my" answer
4. **Assistant information questions**: When testing assistant advice/suggestions, use "What did you suggest/
      recommend/advise..." format
5. **Multi-detail synthesis**: Question should require combining information from different conversation parts
6. **Cross-turn integration**: Force LLM to connect scattered information across multiple turns
7. **Complex reasoning**: Require understanding of relationships and synthesis of multiple elements
8. **Challenging extraction**: Force LLM to demonstrate knowledge through indirect demonstration

Generate ONE high-quality indirect information extraction question that tests recall of specific factual
      details through contextual questioning requiring synthesis of multiple details from different parts of
      the conversation.

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

Listing 10: Information extraction probing question generation prompt

```
You are tasked with generating a probing question to test multi-session reasoning capabilities of LLMs. You
      will be given multiple related bullet points and the corresponding multi-turn dialogs between a user and
       assistant that incorporate this information across different conversation sessions.

Your task is to create ONE question that tests whether an LLM can perform complex multi-hop reasoning,
      synthesis, and analysis across 4+ conversation sessions.

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

## CRITICAL REQUIREMENT: HARD MULTI-SESSION REASONING
- The question MUST NOT include any explicit number, dates, times, duration, or temporal references
- Focus on complex synthesis requiring multi-hop reasoning across 4+ sessions
- Test sophisticated analysis that requires connecting multiple data points
- Ask for complex calculations, patterns, or insights that need advanced reasoning

## QUESTION GENERATION GUIDELINES
Focus on creating questions that require: - **Complex Aggregation** - **Advanced Synthesis** - **Multi-hop
      Reasoning**
- **Pattern Recognition** - **Performance Evaluation** - **Comparative Analysis** - **Predictive Reasoning**

## QUESTION TYPES TO GENERATE (HARD LEVEL)
1. **Complex Multi-hop Calculation** 2. **Performance Evaluation** 3. **Multi-variable Comparison** 4. **
      Complex Evolution Analysis**

## REASONING COMPLEXITY LEVEL: HARD
- **Hard**: Requires complex multi-hop reasoning, synthesis, and analysis across 4+ sessions
- Focus on sophisticated calculations or insights requiring advanced reasoning
- Test ability to identify complex patterns, correlations, or relationships
- Include deep analytical thinking and synthesis of multiple data points

## 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 information needed for reasoning
- List ALL chat_ids where relevant information appears across the reasoning chain
- 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
- If reasoning spans multiple chats, include all relevant chat_ids
```