```
}

## IMPORTANT REQUIREMENTS
1. Do not mention how or when the value changed.Question text must not contain words like 'after,'''negotiated
   ,'''updated,'''revised,'' or any mention of a change process.
2. **Current state focus**: Question must ask about the updated/current information only
3. **No change language**: Avoid words like "changed," "updated," "revised" in the question
4. **Updated answer**: Answer must reflect the most recent version of the information
5. **Confusion potential**: Note what outdated information the LLM might incorrectly recall
6. **Natural phrasing**: Question should sound like asking for current facts, not testing memory updates
7. ''Include at least **two** entries in 'conversation_references': one for the original fact session and one
   for the updated fact session.''

Generate ONE knowledge update question that tests whether the LLM correctly recalls the updated information
   rather than the original outdated version.

CRITICAL NOTE: Do not mention how or when the value changed.Question text must not contain words like 'after,''
   'negotiated,'''updated,'''revised,'' or any mention of a change process.

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

Listing 12: Knowledge update probing question generation prompt

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

Your task is to create ONE question that tests whether an LLM can perform complex multi-step temporal
     reasoning, advanced calculations, pattern analysis, or synthesis of multiple temporal relationships.

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

## CRITICAL REQUIREMENTS: CHALLENGING TEMPORAL REASONING
- The question MUST NOT include any explicit dates, times, or temporal references
- Use only event descriptions that require the LLM to recall temporal information
- Create questions that require complex temporal reasoning, not simple lookups
- Test sophisticated temporal understanding across multiple conversation sessions

## ADVANCED QUESTION GENERATION GUIDELINES
Focus on creating questions that test:
- **Complex duration calculations** - **Relative temporal positioning** - **Cross-session temporal synthesis**
     - **Temporal pattern recognition**
- **Conditional temporal logic** - **Temporal inference**

## SOPHISTICATED QUESTION TYPES
### **Duration & Calculation Questions** 1. **Multi-hop Duration** [Other examples]

### **Sequence & Ordering Questions** 6. **Complex Sequencing** [Other examples]

### **Comparative & Analytical Questions** 10. **Timeline Comparison** [Other examples]

### **Inferential & Complex Questions** 15. **Causal Temporal** [Other examples]

### **Between-Time Information Extraction**:  21. "What/Who/Where/How much/When [specific query] between [
     starting point] and [ending point]?"

## FORBIDDEN QUESTION ELEMENTS
- Do NOT mention specific dates, times, or numbers in the question
- Do NOT use phrases like "on [specific date]" or "after [X] days/weeks/months"
[Other examples]

## GOOD VS BAD EXAMPLES
[Examples]

## TEMPORAL COMPLEXITY LEVEL: HARD
- **Hard**: Requires multi-step temporal reasoning across 3+ conversation sessions, complex calculations,
     pattern analysis, temporal inference, or synthesis of multiple temporal relationships

## 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"
- Make questions sound conversational and natural
- Questions should require deep temporal reasoning to answer

## CHAT ID TRACKING REQUIREMENT
- You MUST identify which specific chat_id(s) contain the temporal information for both events
- List the chat_id for the first temporal event and the chat_id for the second temporal event
- 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:
```