

```
{
  "question": "", "answer": "", "difficulty": "hard", "temporal_type": "", "time_points": "", "
  conversation_references": "",
  "calculation_required": "", "complexity_factors": "", "source_chat_ids": { "first_event": [12, 15], "
  second_event": [28, 31]}
}

## IMPORTANT REQUIREMENTS
1. **No temporal hints**: Question must NOT include dates, times, or temporal references
2. **Complex reasoning**: Require multi-step temporal analysis, not simple lookup
3. **Cross-session dependency**: Events should span multiple conversation sessions
4. **Advanced difficulty**: Always use "hard" difficulty level
5. **Sophisticated types**: Use advanced question types that test complex temporal reasoning
6. **Pattern recognition**: Include questions that test temporal patterns and relationships

Generate ONE high-quality, challenging temporal reasoning question that requires sophisticated temporal
analysis between the provided bullet points.

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

Listing 13: Temporal reasoning probing question generation prompt

```
You are tasked with generating multiple probing questions to test abstention capabilities of LLMs. You will be
given a plan containing bullet points that will be used to generate conversations between a user and AI
assistant.

Your task is to create questions that test whether an LLM can appropriately respond with "I don't know" or "
Based on the provided chat, there is no information related to [topic]" when asked about information
that is missing or unavailable from the plan.

## INPUT DATA
- **PLAN**: <plan>

## CRITICAL REQUIREMENT: EARLY BATCH PRIORITIZATION
**SELECTION PRIORITY ORDER**:
1. **Batch 1-3 (HIGHEST PRIORITY)**: Select 70-80% of your questions from these early batches
2. **Batch 4-6 (MEDIUM PRIORITY)**: Select 10-20% of your questions from these middle batches
3. **Batch 7+ (LOW PRIORITY)**: Select only 5-10% of your questions from later batches

## CRITICAL REQUIREMENT: CONTEXT-BASED HARD QUESTIONS
- Questions MUST be related to the plan's context and domain
- Ask about plausible information that SEEMS like it should be available but isn't
- Use names, places, events, and topics mentioned in the plan
- Ask for specific details about mentioned entities that aren't actually provided
- Create questions that would be natural follow-ups but can't be answered

## QUESTION GENERATION GUIDELINES
Focus on creating questions that test appropriate abstention for:
### 1. Missing Details About Mentioned People/Entities [Examples]
### 2. Unavailable Specifics About Mentioned Events/Activities [Examples]
### 3. Missing Information About Referenced Sources/Materials [Examples]
### 4. Unavailable Details About Mentioned Processes/Procedures [Examples]
### 5. Missing Context About Mentioned Decisions/Choices [Examples]
### 6. Unavailable Quantitative/Measurement Details [Examples]
### 7. Missing Emotional/Subjective Information [Examples]
### 8. Unavailable Technical/Specialized Details [Examples]
### 9. Unavailable Future/Predictive Information [Examples]

## ABSTENTION QUESTION TYPES
Only generate these two types: 1. **Unavailable Information Questions**: Ask about topics, people, events, or
details that are completely absent from the plan
2. **Missing Detail Questions**: Ask for specific details about topics that may be mentioned generally but
lack the requested specifics

## DIFFICULTY LEVELS
Generate questions of varying abstention difficulty:
- **Easy**: Ask for details about mentioned entities that seem like they should be available
- **Medium**: Ask for specific information about mentioned processes/events that weren't detailed
- **Hard**: Ask for nuanced details that are plausible but not provided

## 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

## OUTPUT FORMAT
Return your analysis in this exact JSON format:
[{"question": "", "ideal_response": "", "difficulty": "easy", "abstention_type": "", "why_unanswerable": "",
  "plan_reference": ""},
]

## IMPORTANT REQUIREMENTS
1. **Context-based questions**: Questions must relate to plan content and use mentioned entities
2. **Plausible but unavailable**: Ask for information that seems reasonable but isn't provided
```