

```

- **Expect clarification request**: LLM should ask which contradictory statement is correct
- **Avoid revealing the contradiction**: Question should seem like routine information request

## CRITICAL REQUIREMENT:
- The question MUST be general and broad about the contradicted topic
- Do NOT include specific dates, times, locations, or detailed circumstances
- Do NOT reference specific events or instances that would narrow down the contradiction
- Ask about the general pattern/behavior, not specific occurrences
- This tests whether the LLM can detect contradictions across all related mentions

## QUESTION LANGUAGE REQUIREMENTS
- Write questions as if the USER is asking them naturally
- Use first person ("I", "my", "me") when referring to the user
- Use second person ("you") when addressing the assistant
- 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

## QUESTION FORMAT Generate questions that ask about the contradicted topic directly: [Examples]

## EXPECTED LLM BEHAVIOR The ideal LLM response should: [Example]

## CHAT ID TRACKING REQUIREMENT
- You MUST identify which specific chat_id(s) contain each contradictory statement
- List the chat_id for the first contradictory statement and the chat_id for the second contradictory statement
- 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": "", "ideal_answer": "", "difficulty": "", "contradiction_type": "", "topic_questioned": "", "
  conversation_references": [],
  "tests_for": "", "source_chat_ids": {"first_statement": [ ], "second_statement": [ ]}
}

## IMPORTANT REQUIREMENTS
1. **Natural question phrasing**: Question should sound like normal fact-seeking, not contradiction testing
2. **Topic focus**: Ask directly about the contradicted subject
3. **Contradiction detection expectation**: LLM should notice and ask for clarification
4. **No hint giving**: Don't reveal that there's a contradiction in the question
5. **Clarification seeking**: Ideal response should ask which statement is correct

Generate ONE contradiction resolution question that tests whether the LLM will detect the contradiction and
appropriately request clarification when asked about the contradicted topic.

CRITICAL NOTE: Do NOT include specific dates, times, locations, or detailed circumstances in the question that
make the question easy.
NOTE: Only output the JSON object without any explanation before or after.

```

Listing 17: Contradiction resolution probing question generation prompt

```

You are tasked with generating a probing question to test advanced summarization capabilities of LLMs. You
will be given 6-8 related bullet points about the same topic/theme 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 synthesize and condense complex, multi-
faceted information from across 4+ conversation sessions into sophisticated, comprehensive summaries.

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

## CRITICAL REQUIREMENT: HARD SUMMARIZATION
- Focus on complex information synthesis from 8-10 bullet points
- Test comprehensive synthesis requiring sophisticated analysis
- Require advanced narrative construction with multiple threads
- Ask for summaries that demonstrate deep understanding and integration

## CRITICAL REQUIREMENT: NEUTRAL SUMMARIZATION TESTING
- The question MUST NOT reveal what should be included in the summary
- The question MUST mention *only* the overarching topic-no specific bullet-point details, subtopics, phases,
or technical terms may appear.
- The question MUST NOT hint at the structure or content of the expected answer
- The question should be maximally generic, forcing the LLM to identify and synthesize all relevant
information independently

## QUESTION GENERATION GUIDELINES
Focus on creating questions that test: - **Complex information synthesis** - **Advanced cross-session
condensation** - **Comprehensive overview**
- **Sophisticated narrative coherence** - **Strategic detail prioritization**

## QUESTION TYPES TO GENERATE (HARD LEVEL)
1. **Complex Relationship & Interaction Summary** 2. **Complete Sequence & Event Analysis** 3. **Resource,
Effort, & Timeline Evolution**
4. **Multi-factor Decision Process Review** 5. **Problem-to-Resolution Journey** 6. **Chronological
Development Overview**
7. **Knowledge & Insight Integration Summary** 8. **Complex Negotiation or Agreement Path**

```