

```

## SUMMARIZATION COMPLEXITY LEVEL: HARD
- **Hard**: 8-10 bullet points requiring comprehensive synthesis with detailed progression
- Focus on sophisticated analysis requiring understanding of complex relationships and patterns
- Test ability to synthesize multiple narrative threads and extensive information
- Include advanced narrative elements with multi-layered connections and sophisticated causation

## 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

## CHAT ID TRACKING REQUIREMENT
- You MUST identify which specific chat_id(s) contain the information needed for the summary
- List ALL chat_ids where relevant summary information appears
- 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_summary": "", "difficulty": "hard", "summarization_type": "",
  "bullet_points_covered": , "conversation_sessions": , "key_elements_tested": [], "synthesis_required": "",
  "source_chat_ids": [ , , ...]
}

## IMPORTANT REQUIREMENTS
1. **Comprehensive coverage**: Summary should integrate all key information from 8-10 bullet points
2. **Sophisticated coherence**: Create complex narrative with multiple threads and advanced logical structure
3. **Advanced multi-session synthesis**: Combine information from 4+ conversation sessions
4. **Strategic condensation**: Include extensive important details while maintaining sophisticated narrative structure
5. **Complex question phrasing**: Question should request comprehensive, sophisticated summaries

Generate ONE advanced summarization question that tests the LLM's ability to synthesize 8-10 bullet points into a sophisticated, comprehensive summary.

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

```

Listing 18: Summarization probing question generation prompt

```

This is a plan that contains detailed bullet points about a topic. This plan is used to generate realistic chat conversations between a user and an AI assistant, which are then used to evaluate the long-term memory capabilities of LLMs.

Your task is to analyze this plan and select bullet points that would be most effective for testing instruction following abilities when incorporated into chat conversations.

Analyze this plan and identify bullet points that contain user instructions ideal for testing whether the LLM remembers and follows user-given instructions.

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

Focus on bullet points with:
- **User Instruction** category/label
- **Explicit instruction statements**: "Always", "Never", "When I ask about X, do Y"
- **Behavioral directives**: How the AI should respond or behave
- **Format instructions**: Specific response formats or structures requested
- **Content instructions**: What to include or exclude in responses
- **Process instructions**: How to handle specific types of requests

Look specifically for bullet points labeled as "User Instruction" that contain:
- Clear directive language ("Always provide", "Never include", "When I ask")
- Specific behavioral expectations for the AI assistant
- Conditional instructions ("When I ask about X, do Y")
- Response formatting requirements
- Content inclusion/exclusion rules

Return your analysis in this exact JSON format:
[{"capability": "instruction_following", "batch_numbers": 1, "bullet_numbers": 32,
  "bullet_points": "User Instruction: Always provide detailed cost breakdowns when I ask about budget estimates."}]

Important formatting notes:
- The "batch_numbers" and "bullet_numbers" correspond to each other positionally
- "1" and "32" means: Batch 1 Bullet 32
- Include the full bullet point text as it appears in the plan
- Focus specifically on "User Instruction" labeled bullet points

Select all bullet points labeled as "User Instruction" from each batch (approximately 10 total).

NOTE: Only output the list without any explanation before or after the list.

```

Listing 19: Instruction following probing question generation prompt