

---

```

- **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 9: Candidate selection instruction following prompt**

You are tasked with generating a probing question to test information extraction capabilities of LLMs. You will be given a bullet point and the corresponding multi-turn dialog between a user and assistant that incorporates this bullet point information.

Your task is to create ONE question that tests whether an LLM can precisely extract and recall specific factual details from the conversation through indirect questioning that requires synthesizing multiple details from different parts of the conversation.

```

## INPUT DATA
- **BULLET POINT**: <bullet_point>
- **CONVERSATION TURNS**: <conversation_turns>

## CRITICAL REQUIREMENT: INFORMATION EXTRACTION
- The question MUST NOT directly ask for the information being tested
- Ask about related topics/contexts that require the LLM to synthesize multiple details from different
    conversation parts
- Force the LLM to extract and combine information scattered across different conversation turns
- Make the LLM demonstrate knowledge of facts without being directly asked for them
- Require connecting and integrating information from multiple different conversation elements
- Remove ALL specific details from the question that would give away the answer

## FORBIDDEN QUESTION ELEMENTS
- Do NOT repeat specific names, numbers, or details being tested
- Do NOT mention key characteristics or attributes being extracted
- Do NOT include descriptive words that hint at the answer
- Do NOT reference specific categories or types being tested
- Do NOT use qualifying details that narrow down the answer

## QUESTION LANGUAGE REQUIREMENTS
- Write questions as if the USER is asking them naturally
- Questions MUST ONLY BE from USER language not ASSISTANT
- **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
- Ask about context, or relationships rather than direct facts

## INDIRECT QUESTIONING STRATEGIES
### 1. **Context-Based Recall** Ask about the surrounding circumstances instead of the exact fact
[Example]
### 2. **Comparison Questions** Encourage differentiation between similar elements
[Example]
### 3. **Timeline Integration** Link facts to their sequence in time
[Example]
### 4. **Problem-Solution Context** Frame questions around issues and how they were addressed:
[Example]
### 5. **Discovery and Learning Process** Focus on the origin of knowledge or awareness
[Example]
### 6. **Relationship and Connection Context** Test understanding of associations
[Example]

## FORBIDDEN DIRECT QUESTIONS
[Examples]

## CHAT ID TRACKING REQUIREMENT
- You MUST identify which specific chat_id(s) contain the information being tested - List ALL chat_ids where
    the answer appears

```