
```

- Enable mathematical aggregation across multiple entries (total costs, time durations, quantity counting)
- Allow before/after comparisons of the same entities across different batches
- Require synthesis of information from 3+ different batches
- Create opportunities for complex multi-hop reasoning questions

Return your analysis in this exact JSON format where each object contains multiple related bullet points:
[{"capability": "multi_session_reasoning", "batch_numbers": "1, 2, 3, 5", "bullet_numbers": "10, 4, 7, 7",
 "bullet_points": "Financial & Budget:Cost Estimation: Initial budget set at $12,500, including materials and labor for decoupled framing and MLV. | Financial & Budget:Expense Tracking: Paid $2,200 deposit to QuietFlow Bahrain for HVAC silencing on March 14 via bank transfer. | Financial & Budget:Expense Tracking: Total spent $9,300 by April 1 on materials, labor, and consultant fees. | Financial & Budget:Expense Tracking: Total project spending $12,000 as of May 1, within original $12,500 budget."}
]

Important formatting notes:
- The "batch_numbers" and "bullet_numbers" correspond to each other positionally
- "1, 2, 3, 5" and "10, 4, 7, 7" means: Batch 1 Bullet 10, Batch 2 Bullet 4, Batch 3 Bullet 7, Batch 5 Bullet 7
- Use comma-separated values for batch_number and bullet_number
- Separate multiple bullet_point entries with " | "
- Each group should contain 2-6 related bullet points
- Focus on groups that enable the most sophisticated multi-hop aggregation and comparison questions

Select 8-12 groups of bullet points that would enable the most sophisticated multi-session reasoning questions

.

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

```

Listing 2: Candidate selection multi-hop reasoning 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 PAIRS of related bullet points that would be most effective for testing knowledge update abilities when incorporated into chat conversations.

Analyze this plan and identify bullet points labeled as "Information Update" and match them with their corresponding original facts from earlier in the plan.

```

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

## CRITICAL REQUIREMENT: SPECIAL UPDATE BULLETS WITH ORIGINAL FACTS
Focus on bullet points that:
- **Are labeled "Information Update"**: Look specifically for bullet points with this exact label
- **Have corresponding original facts**: Find the earlier bullet point that contains the original information being updated
- **Show clear before/after relationships**: Original information paired with its explicit update or correction

For each "Information Update" bullet point you find:
1. **Locate the original fact** in an earlier bullet point that this update refers to
2. **Create a pair** with the original bullet point first, then the "Information Update" bullet point second
3. **Ensure clear connection** between the original fact and its update

Look specifically for "Information Update" bullet points that contain:
- Clear update language ("updated," "changed," "revised," "rescheduled," "increased," "decreased")
- References to modifications of previously mentioned information
- Corrections or adjustments to earlier facts
- Timeline or specification changes
Then match each update with its original fact from earlier bullet points.

Return your analysis in this exact JSON format where each object contains exactly TWO related bullet points (original + Information Update):
[{"capability": "knowledge_update", "batch_numbers": "1, 3", "bullet_numbers": "9, 31",
 "bullet_points": ". **Financial & Budget:Cost Estimation:** Initial budget set at $12,500, including materials and labor for decoupled framing and MLV. | Information Update: The initial framing materials purchase included an additional 10% surplus to accommodate unexpected cuts and errors ."}
]

Important formatting notes:
- The "batch_numbers" and "bullet_numbers" correspond to each other positionally
- "1, 3" and "9, 31" means: Batch 1 Bullet 9 (original), Batch 3 Bullet 31 (Information Update)
- Each object must contain exactly 2 bullet points separated by " | "
- Use comma-separated values for batch_number and bullet_number
- First bullet point should represent the original information
- Second bullet point should be the "Information Update" labeled bullet
- Focus on pairs that enable questions like "How did the original plan change when you got the update?"

Select all "Information Update" bullet points and pair them with their corresponding original facts (approximately 10 pairs total).

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

```

Listing 3: Candidate selection knowledge update 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.