G PROMPTS

Here we provide the prompts used in different stages of our framework.

```
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
     information extraction abilities when incorporated into chat conversations.
Analyze this plan and identify bullet points that contain specific factual information ideal for testing
     precise recall and information extraction capabilities.

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

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

Focus on bullet points with:
- **Specific numbers, quantities, measurements, prices, percentages**
- **Proper names of people, organizations, brands, locations**
- **Exact dates, times, schedules, or deadlines**
- **Contact details such as addresses, phone numbers, email IDs**
- **Technical or detailed descriptions (model names, product codes, ratings, specifications)**
- **Distinctive events, awards, or milestones**
- **Direct quotes, messages, or instructions with exact wording**
- **Precise parameters, formulas, or datasets in technical and academic contexts**
- **Mathematical expressions, theorems, proofs**

Prioritize information that:
- Appears early in the timeline
- Contains multiple distinct factual details in one bullet point
- Includes uncommon names, technical terms, or culturally specific references
- Has precise numerical values or measurements that could be easily confused
- Could be misremembered if details are swapped, rounded, or reworded
- Requires high accuracy to preserve meaning (e.g., formulas, addresses, step-by-step processes)

Return your analysis in this exact JSON format:
[{"capability": "information_extraction", "batch_numbers": 1, "bullet_numbers": 3,
        "bullet_points": "• **Personal Introduction:** I am Sherry Rodriguez, 34, licensed conveyancer in
                Hollyborough, Bahrain, earning approximately $68,000 annually."}
]

Important formatting notes:
- The "batch_numbers" and "bullet_numbers" correspond to each other positionally
- "1" and "3" means: Batch 1 Bullet 3

Select ONLY <bullet_number> bullet points total that would generate the highest quality information extraction
        questions.

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

Listing 1: Candidate selection information extraction 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 GROUPS of related bullet points that would be most effective for
     testing multi-session reasoning abilities when incorporated into chat conversations.
Analyze this project plan and identify GROUPS of bullet points that enable testing of aggregation, comparison,
      and synthesis across multiple batches/sessions. Each group should contain 2-6 related bullet points
     that together enable complex multi-hop reasoning questions.

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

## CRITICAL REQUIREMENT: EARLY BATCH PRIORITIZATION
**SELECTION PRIORITY ORDER:**
1. **Groups starting in Batches 1-3 (HIGHEST PRIORITY)**: Select 70-80% of your groups with primary content
     from early batches
2. **Groups spanning early to middle batches (MEDIUM PRIORITY)**: Select 10-20% of groups that bridge early-to
     -middle timeline
3. **Groups from later batches only (LOW PRIORITY)**: Select only 5-10% from purely later batches

Focus on bullet point groups that involve:
- **Aggregation opportunities**: Multiple costs that need to be summed, events that need counting,
     measurements to be totaled
- **Comparison patterns**: Same categories appearing in different batches (budget changes, progress updates,
     relationship interactions)
- **Evolution tracking**: How preferences, decisions, or situations change over time across multiple bullet
     points
- **Cross-reference relationships**: Information that connects between different people, events, or decisions
     across batches

Prioritize bullet point groups that:
- Are part of recurring themes across multiple batches (budget tracking, progress monitoring, relationship
     dynamics)
```