

Appendix A

Prompts to the LLM and the exemplified output are provided in this appendix.

Prompts to LLM and an example of LLM output

To start up the LLM prediction, we first prompt the LLM for role setting, followed by experiment description, constraints, and the request for reasoning. The prompts after tuning are provided as follows.

Prompt for role setting

Imagine you are a chemist conducting a chemical experiment; your task is to predict what elements can be seen in the reaction process. A computer vision system is implemented to fetch real-time vision features of the experiment and update its observation.

Note that in the following description, we will use the word “container” to represent the equipment used in to experiment.

Prompt for experiment description

Description of the experiment you are conducting is:

<start description>

{*Insert experiment description*}

<end description>

Prompt for constraints setting

The elements are defined in 5 classes:

>homogeneous liquid

> heterogeneous liquid

>solid

>residue

>empty

The definition of each class is:

- > homogeneous liquid: a homogeneous liquid is a type of liquid form that could be totally transparent, or with small particles uniformly distributed in the liquid.
- > heterogeneous liquid: a liquid mixture with the particles floating around. The entire layer is not uniform.
- >solid: a solid element could be a large chunk floating in liquid, crystalline in liquid but sinks to the bottom, and a dry chunk at the container bottom.
- >residue: small solid grains stick to the container's window.
- >empty: the air layer on top of the materials when the container is not full.

RULES:

- >Do not create any new class.
- >Do not add explanations after class names in predictions.
- >Residue appears only when a solid sticks to the window, not liquid droplets or invisible particles.
- >Residue and empty will not appear at the same time; just predict one with a higher possibility.
- >The prediction should include elements in different reaction stages, from beginning to end. First, you should analyze how many stages exist in the experiment and predict elements for each stage.
- > The experiment progresses in time, and the timeline of observations changes in the system.
- > It is possible for multiple liquid layers to exist in one container.

Prompt for request reasoning

Please predict what elements will be seen during the experiment with high possibility, and give your response with the following format.

<thought> your thought </thought>

<stage no.>

<start prediction>

>element 1

>element 2

<end prediction>

Example of LLM output

In this example, a crystallization process has been described and input to the LLM. The experiment description

<start description>

An aqueous solution (1.5 mL) of 1 mM BaCl₂ and 2 mM CuSO₄ was stirred continuously in a 4 mL vial.

<end description>

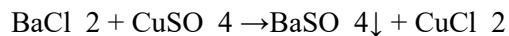
The output from LLM is split into 2 parts:

- <thought> to give an overview of the reasoning process
- <prediction> into different stages.

<thought> is optional for output as the purpose is to follow the reasoning process and track issues for mistaken prediction if necessary.

The output from LLM is:

<thought> The experiment involves mixing two aqueous solutions: BaCl₂ and CuSO₄. When these two salts are combined in aqueous solution, a double displacement reaction can occur:



BaSO₄ is an insoluble salt and will precipitate as a solid. CuCl₂ remains soluble in water, forming a blue-green homogeneous solution.

Given the small volume (1.5 mL total in a 4 mL vial), there will likely be an empty layer at the top. As the reaction proceeds:

Initially, both solutions are clear, homogeneous liquids. Upon mixing, BaSO₄ precipitates, making the solution heterogeneous with suspended solid. With time, BaSO₄ settles to the bottom as a solid layer. If some BaSO₄ adheres to the container wall during stirring or settling, residue may appear.

Since residue and empty are mutually exclusive, and residue is likely during/after precipitation, I will predict residue in the final stage.

The experiment thus has 3 stages:

- Initial clear solution
- Reaction with precipitation and suspension
- Post-reaction with settled solid and possible residue

</thought>

The reasoning result from LLM is:

<stage 1>

<prediction>

homogeneous liquid

empty

</prediction>

<stage 2>

<prediction>

heterogeneous liquid

solid

</prediction>

<stage 3>

<prediction>

solid

residue

</prediction>

Evaluation: In this example, the LLM correctly identified the number of reaction stages, consistent with the manual segmentation of the corresponding video based on phase changes, and most substances at each stage. However, some mispredictions were observed—for

instance, the absence of the solid class in the early phase of Stage 1 and the residue layer in Stage 2. These results indicate that while the LLM can provide an accurate overall interpretation of the experiment, it may occasionally overlook specific transient states. To evaluate the impact of such mispredictions, we conducted a test using randomly selected presented classes as prompts and the result is shown in Table 3. In the current framework, YOLO-text receives LLM-predicted stage classes as prompts to perform detection on corresponding video subsets. This process highlights the need for a dynamic prompting system capable of automatically updating the prompt classes as the experiment progresses, thereby enhancing adaptability to longer or multi-stage reactions.

Appendix B

To evaluate the LLM's performance with prompts phrased in different styles, several prompt variations were generated based on the baseline experiment description used in the previous cases. Variation 1 removed detailed descriptions of the beginning, reaction, and completion phases. Variation 3 rephrased the description in a more formal tone with a specific class given, while Variation 2 adapted it to a scientific report style. The results show that the LLM generally produces consistent predictions of stages and classes across these variations under the guidance of the applied constraints. However, in Variation 3, if a pre-defined class e.g. homogeneous mixer of oil and water, was explicitly mentioned in the experiment description, the LLM tended to rely on that reference and predict the class according to the given information.

Baseline prompt

<start description>

This is a liquid-liquid extraction type experiment where I will conduct the extraction of oil and water. The phases included in the entire reaction process are described as followed. Beginning: A mixture of oil and water. During reaction: The oil will gradually separate from the water and float to the surface over time. Complete: The oil and water are separated distinctly.

<end description>

The reasoning output from the LLM is

<thought> In this liquid-liquid extraction experiment, I am working with a mixture of oil and water. The mixture of oil and water is immiscible, meaning they will form two distinct layers due to differences in density and polarity. The experiment will go through phases where the oil and water separate due to their immiscibility. The process will start with a heterogeneous mixture, leading to eventual separation into layers. There may also be an air gap at the top if the container is not completely filled. Given the nature of the materials and the extraction process, I expect the phases to show different types of liquid forms and an air gap at the end.

</thought>

<phase 1>

<start prediction>

heterogeneous liquid

empty

<end prediction>

<phase 2>

<start prediction>

heterogeneous liquid

homogeneous liquid

empty

<end prediction>

<phase 3>

<start prediction>

homogeneous liquid

homogeneous liquid

empty

<end prediction>

The reasoning output gives comprehensive analysis and correct class prediction according to the given definition. The output from prompt variations will be provided in the next step to study how wording will affect the LLM performance by comparing with the baseline.

Variation 1: General description

<start description> This is a liquid-liquid extraction type experiment where I will conduct the extraction of oil and water. The phases included in the entire reaction process are described as follows. <end description>

The LLM provides same predictions as baseline.

Variation 2: Prompt as academic report

<start description> In this liquid-liquid extraction experiment, the separation behavior of an oil-water system is investigated. Initially, the system consists of a dispersed mixture of oil and water. As the extraction progresses, the oil phase begins to coalesce and migrate upward, forming a distinct upper layer. Upon completion, the two immiscible phases—oil and water—are clearly separated, indicating successful phase partitioning.<end description>

The LLM provides same predictions as baseline.

Variation 3: Prompt with specific class given

In this example, the experiment description is rephrased in a formal style. However, the oil-water mixer was described as homogeneous liquid in the given script.

<start description> This experiment involves a liquid–liquid extraction process focusing on the separation of oil and water. The procedure progresses through the following stages: Beginning: A homogeneous mixture of oil and water. During reaction: The oil gradually separates from the water and rises to the surface. Completion: The two phases become distinctly separated, forming clear oil and water layers.<end description>

Output from LLM is:

<thought> The experiment follows a liquid–liquid extraction process involving oil and water. Based on the description, the experiment progresses through three stages:

Beginning – A homogeneous mixture of oil and water implies that they may be temporarily emulsified or thoroughly mixed, likely forming a single-phase appearance or a uniform dispersion.

During reaction – The oil separates and rises, indicating the emergence of two distinct liquid layers (heterogeneous).

Completion – Clear separation into distinct oil and water layers.

We should expect different element classes at each stage, transitioning from homogeneous liquid to heterogeneous liquid, and then distinct layers of homogeneous liquids. Empty may be present if there's space at the top of the container. Solid and residue are not relevant as no solids are involved in the described experiment.

</thought>

<stage 1>

<start prediction>

homogeneous liquid

empty

<end prediction>

<stage 2>

<start prediction>

heterogeneous liquid

empty

<end prediction>

<stage 3>

<start prediction>

homogeneous liquid

homogeneous liquid

empty

<end prediction>