

# EfficientLLaVA: Generalizable Auto-Pruning for Large Vision-language Models

## Supplementary Material

### A. Visual Answer



Is Nerium oleander made up of many cells?

**LLM-Pruner:**  
Yes. Some phyla have many different types of organisms. Other phyla have only one member. Nerium oleander is a plant. Plants are made up of many cells.

**SparseGPT:**  
How many cells do they have? one | one | one or many | one or many | many | many

**EfficientLLaVA:**  
Yes, Nerium oleander is a plant. Plant cells have many more cell types than most animal cells. In a plant, each cell does a different job. Some cells make sugar that the plant can use as food. Other cells make the substances that help the plant grow and survive.



Which property do these four objects have in common?

**LLM-Pruner:**  
The tape and the melted marshmallow are not sticky and stretchy. The property that all four objects have in common is slippery.

**SparseGPT:**  
The tape is stretchy, but the glass bottle is not. The property that all four objects have in common is slippery.

**EfficientLLaVA:**  
A stretchy object gets longer when you pull on it. The chocolate milkshake is not stretchy. A sticky object can attach or stick to other things. All four objects are sticky. A slippery object is hard to hold onto or stand on. The tape is not slippery. The property that all four objects have in common is sticky.

Figure 1. Visual examples from LLaVA-SQA-7B. We color the text to show the response of different pruning methods and EfficientLLaVA consistently delivers more refined, contextually appropriate responses, showcasing its superior pruning and reasoning capabilities.