

## **TRAINING**

## Stage 1: Multimodal Alignment Tuning



## Stage 2: Multimodal Instruction Tuning



A deer strides through snowy landscape, highlighted by

the cold air.

cup of coffee