

SmolVLM-500M-Instruct

| Introduction

SmolVLM-500M is a tiny multimodal model, member of the SmolVLM family. It accepts arbitrary sequences of image and text inputs to produce text outputs. It's designed for efficiency. SmolVLM can answer questions about images, describe visual content, or transcribe text. Its lightweight architecture makes it suitable for on-device applications while maintaining strong performance on multimodal tasks. It can run inference on one image with 1.23GB of GPU RAM.

| Available NPU Models

| Base Model

~~~~~

smolvlm-500M-ax630c ^^^^^^^^^^^^^^^^^

The **Base Model** providing a 128 context window and a maximum output of 1,024 tokens.

**Support Platforms:** LLM630 Compute Kit, Module LLM, and Module LLM Kit

- 128 context window
- 1,024 max output tokens
- ttft 365.69ms
- avg-token/s 13.14
- Image encoding size 512\*512
- Image encoding time 838.30ms

## | Install

\*\*\*\*\*

```
apt install llm-model-smolvlm-500m-ax630c
```

**Manual installation:** [Click here to download llm-model-smolvlm-500m-ax630c](#)