

Based on the LeRobot documentation and the specific architecture of SmolVLA, here are the answers to your setup questions.

## 1. Do I have to do Diffusion before testing SmolVLA?

**No, you do not.** You can skip the "Diffusion Policy" step and go straight to fine-tuning SmolVLA.

- **Why the confusion:** The previous report suggested a "Phased Approach" (Phase 1: Diffusion, Phase 2: VLA) purely as a *debugging strategy* to test your data pipeline quickly. It is **not** a software requirement.
- **How it works:** SmolVLA uses a technique called **Flow Matching** (which is mathematically similar to diffusion) to generate actions. When you run the LeRobot fine-tuning script, it trains this Flow Matching "action head" directly. You do not need a separate diffusion model as a prerequisite.
- **Recommendation:** Since you want to use LeRobot and training is relatively efficient, you should feel comfortable training **SmolVLA directly** as your first model.

## 2. Can I start by capturing a very simple "Drive from Point A to Point B"?

**Yes, but you must define "Point B" visually.**

VLA models are "goal-conditioned" or "instruction-conditioned." The robot needs to know *what* Point B is.

- **Bad Task:** "Drive forward 2 meters." (Hard for a VLA to learn because "forward" looks different depending on where you are standing).
- **Good Task:** "Drive to the red box."
- **Why:** This gives the visual encoder a target to lock onto. The model learns: *"When I see the red box is small/far away, output high velocity. When the red box is large/close, output zero velocity (stop)."*

### Capture Strategy:

Place a distinct object (e.g., a colored box or cone) as your target "Point B" and record the robot driving to it.

## 3. How many records should I have?

### Aim for 50 episodes.

According to LeRobot's specific guidelines for fine-tuning SmolVLA:

- **Minimum Viable: 50 episodes** is the recommended starting point. Experiments showed that 25 episodes were insufficient for reliable performance, while 50 episodes provided good generalization for basic tasks.
- **Structure:** Do not record the exact same drive 50 times. You need **diversity** to prevent overfitting.
  - **Variation Strategy:** Break the 50 episodes into 5 variations (10 episodes each).
    1. Start robot 1 meter away, straight on.

2. Start robot 1 meter away, offset to the left.
  3. Start robot 1 meter away, offset to the right.
  4. Start robot 1.5 meters away.
  5. Change the background slightly (e.g., move a chair in the room).
- **Duration:** Keep episodes short and focused (e.g., 10-15 seconds).

### **Summary Checklist for First Capture**

- ☐ **Task:** "Drive to" (ensure target is visible in Static & Onboard cameras).
- ☐ **Model:** SmolVLA (train directly).
- ☐ **Volume:** 50 Episodes total.
- ☐ **Diversity:** 5 different starting positions/angles relative to the target.