

Leader-follower vs direct teleoperation

Notes from a chat with Claude, October 3–4, 2026.

Question: Why do you need leader-follower teleoperation? Why not teleoperate the robot directly?

Leader–follower is direct teleop; the leader is just the input device. The robot needs a target for every joint about 50 times a second, so every teleop setup needs something that turns your intent into those numbers. The question is which device does that best. There are three common options:

1. Joystick, gamepad or 3D mouse. You steer one point, the gripper, by velocity ("move up, rotate left"). Software then works out the 6–7 joint angles; this calculation is called inverse kinematics.

- You're controlling six axes with two thumbs, so you work in slow steps: move, then rotate, then close.
- The software picks the elbow position. Near certain poses or joint limits, the arm stalls or flips.
- Two arms at once is nearly impossible.
- The result is slow, robotic demos, and the policy copies them.

2. VR or hand tracking. You move your real hands, and tracking maps them onto the robot. It's fast and natural, but:

- Your arm isn't the robot's arm, so some of your poses are unreachable or mapped awkwardly.
- Tracking jitters and drops out when your hands are hidden from the cameras.
- Nothing stops you: your hand passes through the table, while the robot's gripper hits it.

3. Leader arm. This is a cheap, often scaled-down copy of the robot with a sensor in every joint. Each joint angle is copied to the matching robot joint. Angles don't depend on size, which is why a small 3D-printed GELLO can drive a full-size arm. That one choice fixes the problems above:

- **No calculation in between.** Copying joint to joint means nothing can flip or stall, and you set the elbow yourself.
- **Every demo is one the robot can do.** The leader has the robot's shape and joint limits, so you physically can't demonstrate a pose the robot can't reach.
- **Fast and intuitive.** You just move an arm. Two hands work naturally for two-arm tasks, a trigger works the gripper, and the delay is negligible.
- **You feel the geometry.** The leader's shape, its weight, and resting it on the table give you a sense of the workspace that a joystick never does.

If by "directly" you mean grabbing the robot and guiding it by hand (called **kinesthetic teaching**), there's a bigger problem. Your hands and body appear in the robot's camera images, so the policy learns from scenes that will never occur when it works alone. Heavy arms are also hard to move smoothly, and it's awkward to work the gripper at the same time. Leader–follower keeps the human out of the robot's cameras.

The leader arm has costs too:

- You need one built for each robot type.

- You feel no contact force, unless the leader has motors that push back ("bilateral" teleop).
- It doesn't extend to a whole humanoid body.

That last point is worth thinking through. Your G1 plan uses a VR headset with NVIDIA's SONIC controller instead of leader arms. Why do you think humanoid teams mostly use VR or motion-capture suits for the whole body? And what would you expect those demos to do worse than ALOHA-style leader-arm data?