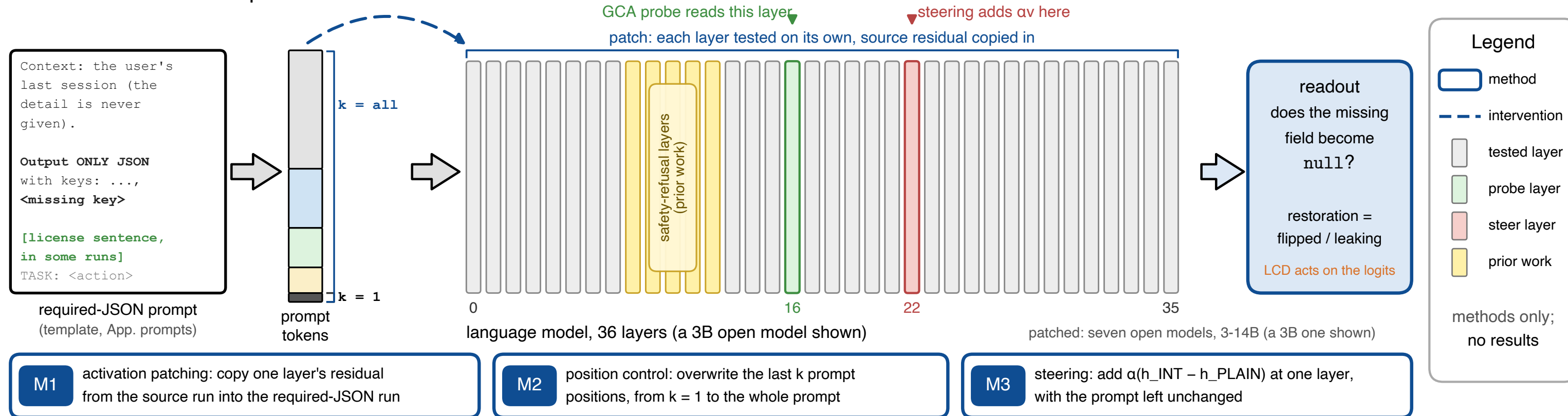


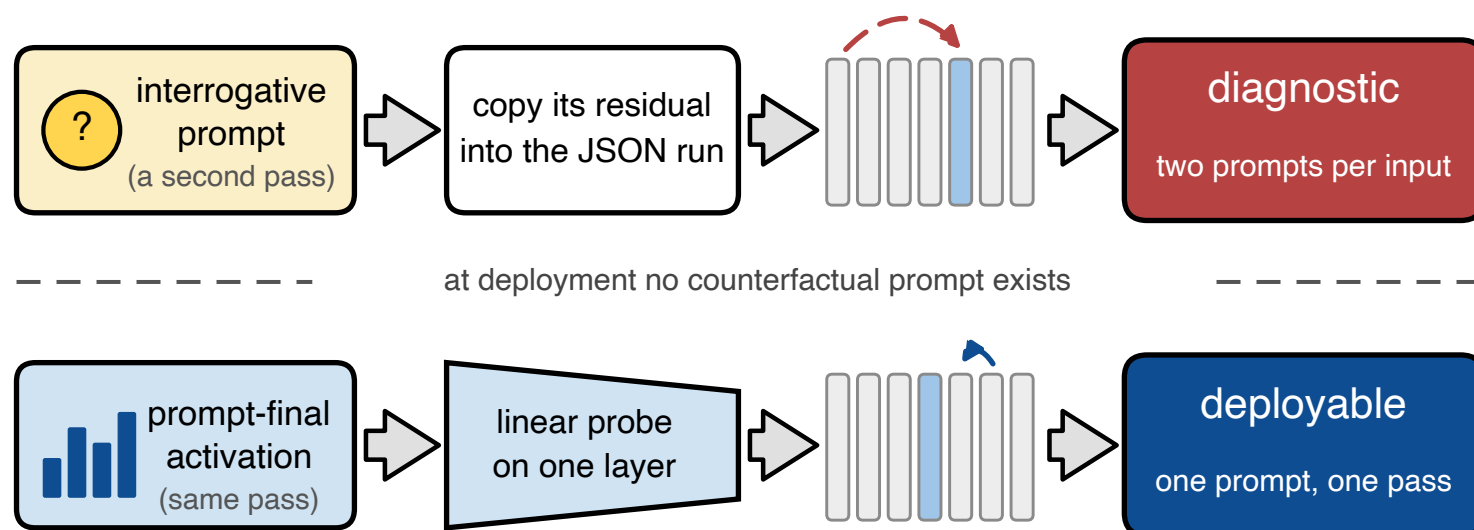
a

How we look inside an open model: what each intervention touches



b

The patch needs a second prompt; the probe does not



c

How one patch is run

