Avg Indirect Effect of $h_i^{(l)}$ over 7 prompts [occupation] [occupation] Question -Begining of context -First subject token -Middle subject tokens Last subject token -Context in between tokens -First object token -Middle object tokens -Last object token -Rest of context tokens -Last token 9 10 11 AIF 8 single patched layer