

|                 | 0.71 | 0.71 | 0.71 | 0.68 | 0.71 |
|-----------------|------|------|------|------|------|
| <b>Original</b> | 0.71 | 0.71 | 0.71 | 0.68 | 0.71 |
| <b>GradDiff</b> | 0.28 | 0.65 | 0.53 | 0.31 | 0.41 |
| <b>NPO</b>      | 0.27 | 0.65 | 0.50 | 0.28 | 0.41 |
| <b>SimNPO</b>   | 0.28 | 0.64 | 0.55 | 0.30 | 0.50 |
| <b>RMU</b>      | 0.27 | 0.70 | 0.41 | 0.26 | 0.59 |
| <b>RR</b>       | 0.32 | 0.71 | 0.45 | 0.36 | 0.51 |
| <b>ELM</b>      | 0.32 | 0.70 | 0.52 | 0.32 | 0.68 |
| <b>DPO</b>      | 0.33 | 0.66 | 0.34 | 0.28 | 0.38 |
| <b>IDK+AP</b>   | 0.32 | 0.52 | 0.40 | 0.29 | 0.29 |
| <b>NPO+SAM*</b> | 0.26 | 0.51 | 0.33 | 0.26 | 0.28 |
| <b>NPO+IRM*</b> | 0.27 | 0.56 | 0.34 | 0.26 | 0.30 |
| <b>RMU+LAT*</b> | 0.31 | 0.71 | 0.43 | 0.26 | 0.57 |
| <b>TAR*</b>     | 0.27 | 0.68 | 0.33 | 0.25 | 0.29 |

Unlearned      Rob<sup>ReL</sup>      Rob<sup>FT</sup>      Rob<sup>QT</sup>      Rob<sup>JA</sup>