| Language-wise Neuron Overlap Rate on bloom-1b1, Layer 17 | | | | | | | | | | | | |
|---|---------|------|------|------|------|------|------|------|------|--|---|--------|
| on broom-rot, hayer it | | | | | | | | | | | | |
| | ar | 0 | 0.04 | 0.06 | 0.06 | 0.08 | 0 | 0.06 | 0.12 | | | |
| Target Languages | es - | 0.06 | 0.12 | 0.1 | 0.16 | 0.1 | 0 | 0.2 | 0.28 | | - | 0.25 |
| | en | 0.02 | 0.1 | 0.06 | 0.06 | 0.06 | 0 | 0.08 | 0.1 | | | |
| | fr - | 0.06 | 0.12 | 0.08 | 0.18 | 0.22 | 0 | 0.24 | 0.26 | | - | 0.20 |
| | id - | 0.02 | 0.08 | 0.02 | 0.1 | 0.06 | 0.08 | 0.06 | 0.1 | | | |
| | mr | 0 | 0.02 | 0.02 | 0.02 | 0.06 | 0.02 | 0 | 0.04 | | | - 0.15 |
| | pt | 0.1 | 0.08 | 0.12 | 0.26 | 0.2 | 0.06 | 0.14 | 0.24 | | | 0.10 |
| | ta - | 0.02 | 0.02 | 0.06 | 0.06 | 0.1 | 0.02 | 0.06 | 0.14 | | | 0.10 |
| | ur - | 0.02 | 0 | 0.02 | 0.02 | 0.06 | 0.04 | 0.06 | 0.14 | | | |

0.12

0.1

300

0.06

0.08

1

0.08

0.08

10

0.04

0.06

100

0.1

0.12

200

Global Steps (K)

0.02

0.02

400

0.1

0.16

500

0.12

0.18

600

0.05

0.00