Attention Scaling

Legend:
- Yarn, full layers
- 8M theta, full layers
- 8M theta, all layers
- 500k theta

X-axis: RULER Average (40, 50, 60, 70, 80, 90, 100)

Y-axis categories: 65K, 32K, 16K, 8K, 4K