

The graph displays the performance of two methods, Baseline and Relative, across a range of token counts from 0B to 10B. The Y-axis represents a performance metric, with values ranging from 0.0 to 1.0. The X-axis is labeled 'Tokens' and has major ticks at 0B, 0B, 0B, 1B, 1B, 1B, 1B, and 10B. The legend indicates four data series: Baseline - mean (blue line), Baseline - standard deviation (light blue shaded area), Relative - mean (red line), and Relative - standard deviation (light red shaded area). Both methods show a decreasing trend in performance as the number of tokens increases. The Baseline method consistently outperforms the Relative method, with a higher mean performance and a smaller standard deviation across the entire range of token counts.

Tokens	Baseline - mean	Baseline - standard deviation	Relative - mean	Relative - standard deviation
0B	0.95	0.05	0.90	0.05
0B	0.85	0.05	0.80	0.05
0B	0.75	0.05	0.70	0.05
1B	0.65	0.05	0.60	0.05
1B	0.55	0.05	0.50	0.05
1B	0.45	0.05	0.40	0.05
1B	0.35	0.05	0.30	0.05
10B	0.25	0.05	0.20	0.05

