Natural language query: A jockey in red and white riding a horse.

Binary mask: $x_1$=253, $y_1$=2, $x_2$=123, $y_2$=23, ... $x_N$=376, $y_N$=23.

Bounding box: $x_1$=87, $y_1$=2, $x_2$=500, $y_2$=492.