

# Percentage of tokens in the top web domains of C4 by robots.txt restriction category, 2016–24

Source: Longpre et al., 2025 | Chart: 2025 AI Index report

